UNIVERSITA' DEGLI STUDI DI TRIESTE

XXVI CICLO DI DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE

Kernel-Based Continuous-Time Systems Identification: Methods and Tools

Settore scientifico-disciplinare **09/G1 - AUTOMATICA**

DOTTORANDO

Andrea Assalone

COORDINATORE DEL COLLEGIO DEI DOCENTI

Chiar.mo Prof. Roberto Vescovo

SUPERVISORE DI TESI

Chiar.mo Prof. Thomas Parisini

ANNO ACCADEMICO 2012/2013

Kernel-Based Continuous-Time Systems Identification: Methods and Tools

By

Andrea Assalone

Dissertation

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in Information Engineering

at Department of Engineering and Architecture

Trieste University, 2014

Trieste, Friuli Venezia Giulia, Italy

Abstract

This thesis is aimed at the formalization of a new theoretical framework, arising from the algebra of Fredholm-Volterra linear integral operators acting on Hilbert spaces, for the synthesis of non-asymptotic state and parameter estimators for continuous-time dynamical systems from input-output measurements subject to time-varying perturbations.
In order to achieve non-asymptotic estimates of continuous-time dynamical systems, classical methods usually augment the vector of decision variables with the unknown initial conditions of the non measured states. However, this comes at the price of an increase of complexity for the algorithm. Recently, several algebraic estimation methods have been developed, arising from an algebraic setting rather than from a statistical or a systems-theoretic perspective. While the strong theoretical foundations and the non-asymptotic convergence property represent oustanding features of these methods, the major drawback is that the practical implementation ends up with an internally unstable dynamic. Therefore, the design of estimation methods for these kind of systems is an important and emergent topic.
The goal of this work is to present some recent results, considering different frameworks and facing some of the issues emerging when dealing with the design of identification algorithms. The target is to develop a comprehensive estimation architecture with fast convergence properties and internally stable.
Following a logical order, first of all we design the identification algorithm by proposing a novel kernel-based architecture, by means of the algebra of Fredholm-Volterra linear integral operators.
Besides, the proposed methodology is addressed in order to design estimators with very fast convergence properties for continuous-time dynamic systems characterized by bounded relative degree and possibly affected by structured perturbations. More specifically, the design of suitable kernels of non-anticipative linear integral operators gives rise to estimators characterized by convergence properties ideally "non-asymptotic".
The analysis of the properties of the kernels guaranteeing such a fast con-

vergence is addressed and two classes of admissible kernel functions are introduced: one for the parameter estimation problem and one for the state estimation problem. The operators induced by the proposed kernels admit implementable (*i.e.*, finite-dimensional and internally stable) state-space realizations.

For the sake of completeness, the bias analysis of the proposed estimator is addressed, deriving the asymptotic properties of the identification algorithm and demonstrating that the kernel functions can be designed taking in account the results obtained with this analysis.

Table of Contents

List of Figures

Chapter 1

Introduction

Modern systems are getting more and more complex and many engineering applications need a compact and accurate description of the dynamic behavior of the system under consideration: this is especially true in the field of automatic control applications. Dynamic models describing the system of interest can be constructured using the first principles of physics, chemistry, biology, and so forth. However, models constructured in this way are difficult to derive, because they require detailed specialist knowledge, which may be lacking, therefore the resulting models are often very complex. Developing in such way models can be very labor-intensive, and hence expensive. For poorly understood systems, the derivation of a model from the first principles is even impossible: since the first-principles models are often complex, their simulation takes considerable time on a computer; therefore, they are not suitable for fast on-line applications for example. Moreover, these models are not always very accurate, because of many reasons, but the most important are the following two: first, it is difficult to decide which effects are relevant and must be included in the model, and which effects can be negligible; second, certain quantities needed to build the model are unknown, and have to be estimated by performing dedicated experiments. The resulting estimates often differ from the real quantities and hence some model mismatch can occur.

An alternative way of building models is through system identification. System identification is the process of developing or improving the mathematical representation of a physical system using observed data: in general,

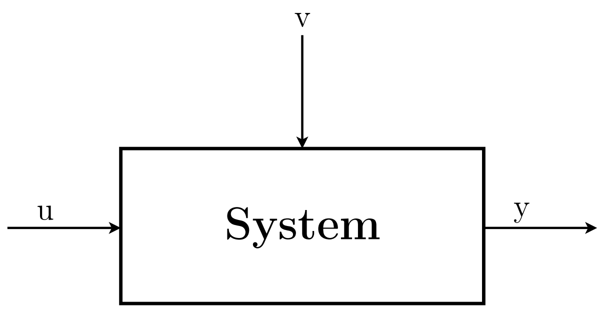


Fig. 1.1: Identification problem

this problem can be seen as in Figure 1.1, whence a dynamic system is considered with input u, output y and disturbance v; we can observe u and y but not v; we can directly manipulate the input u but not y. Even if we do not know the inside structure of the system, the measured input and output data provide useful informations about the system behavior; thus we can construct mathematical models to describe dynamics of the system of interest from observed input-output data. Therefore, first principles are not directly used to model the system, but expert knowledge about the system still plays an important and key role. Such knowledge is of great value for setting up identification experiments to generate the required measurements, for deciding upon the type of models to be used and for determining the quality and validity of the estimated models. System identification often yelds compact accurate models that are suitable for fast on-line applications and for model-based predictive control, which has found widespread use in several branches of the industrial processes. System identification still requires considerable human invention and expert knowledge to obtain models that are satisfactory for the application in mind. Nevertheless, compared with the development of models from first principles, it is not so labor-intensive. At present, several steps of the identification procedure can be automated.
It is important to remark that building models from first principles only and system identification based only on measurements are two extreme cases. Quite often a combination of the two is encountered. In such a combination, system identification is used to estimate the unknown parts or parameters of a model based on first principles. Therefore, a whole range of approaches exists that shows a gradual transition from first-principles modeling to system identification.
We can look at the identification problem, using the definition [1]:
"Identification is the determination, on the basis of input and output, of a system within a specified class of systems, to which the system under test is equivalent."
Using Zadeh's formulation it is necessary to specify a class of systems, a class of input signals, and the meaning of "equivalent". Therefore, in general, system identification proceeds as follows: first, a certain type of model is selected that is considered to be suitable for the applicant at hand. The model class is the set of systems that can be represented by a certain model structure; the choice of model class requires some knowledge of the system under observation: if the system is not within the model class (or sufficiently close to it), the estimated model will not be useful. Different classes of models can be used such as the Oe (Output Error), ARMAX (Autoregressive Moving Average eXogenous), etc.. Also the representation has to be determined, possible choices include the state space model, a frequency response function or a transfer function model: transfer function

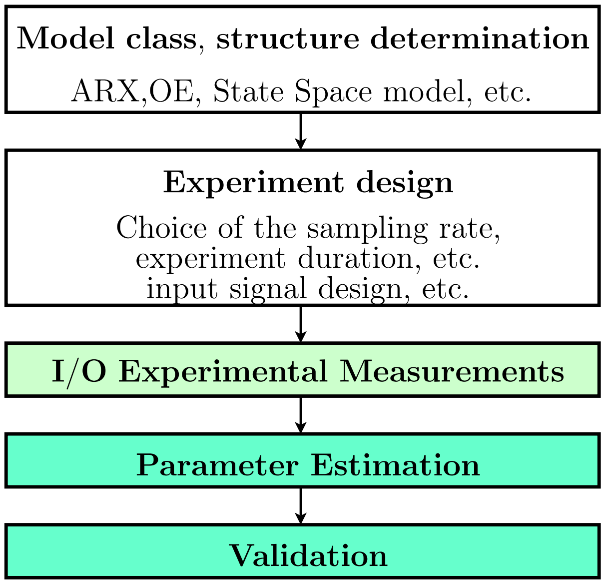


Fig. 1.2: Diagram of system identification process

models are most common due to their low number of parameters. Second, a special input signal is designed such that the model captures the behavior of the system to be modeled; third, an identification experiment is carried out in which input and output signals are measured; in this part we include the pre-processing of the data measured (filtering, detrending, decimating, etc.). An identification method is then selected to estimate the parameters that describe the model from the collected input and output measurements; finally, the validity of the obtained model is evaluated (the flow chart representation of the system identification procedure is depicted in figure 1.2).

An important step in system identification is the determination of the type of model to be used. This decision is based on knowledge of the system under consideration, and on the properties of the model. Certain types of models can be used to approximate the input-output behavior of a smooth nonlinear dynamical system up to arbitrary accuracy. Such models have the so-called universal approximation capability. An example of a universal approximator are the neural networks [2]. The drawback of these models is that they are complex, and difficult to estimate and analyze. Therefore other models structures have received considerable attention over the years.

1.1 Structure and contributions

In this work, the results presented in [3], [4] and [5], where a kernel-based system identification for continuous-time linear systems is designed, are extended in orther to face some of the issues emerging when a noise with stochastical structure arise. The objective is to prove that the synthesis of an internally stable dynamic non-asymptotic estimator can be carried out by devising the kernel of a non-anticipative linear integral operator, yelding to a minimal nonlinear dynamic implementation.
Many problems of mathematical physics, theory of elasticity, viscodynamic fluids and others reduce to Fredholm-Volterra integral equations whence the kernel function could takes several special forms: the characterization of the class of admissible kernels is of great interest, since the solution of integral equations depends on the properties that those kernels have. The use of kernels in our framework is aimed at obtaining implementable (non-anticipative) dynamic filters with minimal realization, that is, in which only the variables of interest appear explicetely. In conclusion, the use of a kernel-based approach is then well suited for the synthesis of non-asymptotic state and parameter estimators of continuous-time dynamical systems from input-output measurements.
Indeed, by resorting to the algebra of linear integral operators, the effects of unknown initial conditions on the states of the observed system could in principle be eliminated from structural equations by a proper choice of the kernel functions. In this setting, the proposed methodology aims first at identifying a class of admissible kernels (*i.e.*, those yielding to non-anticipative internally stable non-asymptotic estimators), establishing the relation with current literature, and finally at characterizing the properties of specific kernel functions.

Summing up, the main idea is to use the algebra of linear integral operators to design kernel functions such that the influence of initial conditions is removed from the estimates. The iterated application of operators induced by those kernels on both sides of the structural constraint of the unknown system will allow to overcome the unavailability of signal derivatives (hidden internal states of the system).
Moreover we consider the following aspects that can augment the performance of a system identification method:

- unified algebraic synthesis: to provide a comprehensive procedure for both parametric and state estimators, it will be established common patterns toward the formulation of a unified algebraic synthesis method;

- internal stability of the system: to avoid problems like wind-up of the integrators and therefore periodical re-initialization of the identification procedure, it will be characterized a class of admissible kernels, in order to have a implementable (non-anticipative) internally stable estimator;

- it will be formalized design methods, that is, introducing optimally criteria for choosing one among admissible kernels

The first point is addressed introducing linear integral operators' algebra, that represents the theoretical basis underlying the present thesis. We will consider the algebra of Fredholm-Volterrra linear integral operators, acting on Hilbert spaces ([6]). This fact allows to consider a wide class of operators with some strong relevant properties in terms of stability and convergence time.
The second point will be tackled considering the properties of Volterra operators: in particular, it will be shown that the stability of the Volterra operators induced by the kernels reflects on the stability properties of the estimator.
The last point will be addressed analyzing the stochastic properties of the proposed estimator and deriving rules for the design of the kernel function.

The thesis is organized as follows: in Chapter 2 it will be presented the theoretical foundations of the proposed methodology and it will be provided the tools for the further discussion. Later, in Part I, the proposed methodology for parametric estimation is described in detail:

- Chapter 3 tackles the possible solutions for the kernel functions design and its properties, in terms of stability and convergence time;

- Chapter 4 the simulation results are presented and discussed

In Part II it will be presented the kernel-based state estimation methodology, in particular:

- Chapter 5 introduces a new possible instance of the kernel function for observers design, that yields to an internal stable linear system

- Chapter 6 discusses and derives the bias analysis for the identification technique proposed; it will be shown the dependency for the kernel parameter with the sampling interval

- Chapter 7 the simulation results are presented and discussed with respect to the observer design

Finally in Chapter 8 some concluding remarks are given and future developments are discussed.

1.2 The state of art

Identification of continuous-time (CT) linear time-invariant (LTI) models for continuous-time dynamic processes was the initial goal in the earliest work on sistem identification. This engineering filed has begun to blossom in the 60's with the works of Aström ([7], [8]), Bohlin ([9], [10]), Ho and Kalman [11] Box and Jenkins [12], Eykhoff [13], Ljung [14] (and many others, see e.g. [15] and [16]). Therefore we can state that the field of system identification has its real roots in the 1960's. The field has expanded enormously in the last 50 years; system identification has become a popular field of research and many articles and books have been published on the topic. This led to rapid advances in the theoretical foundation as well as in the practical application of system identidication. However, this large variety of methods has made it difficult for the user to choose the best suited method for the problem at hand.
The LTI models has been used succesfully in many engineering applications, and a considerable body of theory exists for system identification and automatic control of linear systems. The authoritative guide for linear system identification is the book by Ljung ([14]); attractive methods for linear systems identification are the subspace methods developed ([17], [18], [19] and [20]): these are numerically robust methods that can easily deal with systems having myltiple inputs and outputs and are noniterative, unlike many other identification methods. Subspace methods were originally developed for the identification of linear systems and are based on numerical methods from linear algebra.
However, due to the developments in the digital data acquisition and computing technology and the concomitant sampled data, led to an emphasis on the use of discrete-time (DT) system models, discrete-time control designs and discrete-time-based system identification algorithms from the mid 1960s onward. The last decade has, however, witnessed a renewed interest in the techniques for the identification of CT models from sampled data and only early research on system identification focused on identification of CT models from CT data. Well extablished theories have been devoloped [16], [14] and many applications have been reported.
A simplistic way of estimating the parameters of a CT model by an indirect approach is to use the sampled data to first estimate a DT model and then convert it into an equivalent CT model. The difficulties to convert a discrete-time transfer function to continuous-time transfer function are well known and related to the accurate estimation of the zeros of a CT

transfer function for non-uniform sampling ([21],[22]).
Direct continuous-time model identification and estimation is advantegeous for a number of reasons that have pratical importance: first, most scientific laws used in scientific model formulation (mass and energy conservation, gravitational laws) are more naturally formulated in continuous-time differential equation terms; second, while discret-time models have different parameter values, depending upon the sampling interval of the data, CT models are defined by a unique set of parameters that are independent of the sampling interval. Third, in case of irregularly sampled data, there are several CT methods that can be easily adapted to (see [23]). Finally, perhaps most importantly, CT models can be identified and estimated from rapidly sampled data, whereas DT models encounter difficulties when the sampling frequency is too high in relation to the dominant frequencies of the system under study ([24]): in this situation, the eigenvalues lie too close to the unit circle in the complex domain and the DT model parameter estimates become statistically ill-defined. The practical consequence of this situation, are either that the DT estimation fails to converge properly, providing an erroneous explanation of the data: or, even if the convergence is achieved, the CT model, as obtained by standard conversion from the estimated DT model, does not provide the correct CT model.
Another important issue whence system identification play a key role is, in control system design, to obtain an accurate model of the plant to be controlled. Though most of the existing identification methods are described in discrete-time, it would often be convenient to have continuous-time models directly from sampled I/O data. Indeed many controller design approaches are cast in a continuous-time set-up and, moreover, it is often easier for us to capture the plant dynamics intuitively in continuous-time rather than in discrete-time. However, a basic difficulty of continuous-time identification is that standard approaches (at times called *standard methods*) require to compute the time-derivatives of I/O data, a nontrivial and very delicate task in the presence of measurement noise. A comprehensive survey and of the attempts made to overcome it, has been first given by [25] and [26]. For more information on direct methods, the reader is referred to the book [27]. Furthermore, the Continuous-Time System Identification (CONTSID) toolbox has been developed on the basis of these direct methods [28][29] and [30].
In this context, it is important to cite iterative learning control (ILC) that has attracted much attention over last two decades even in the identification of continuous-time systems. Indeed this identification method provide several advantages such as: 1) no time-derivatives I/O data are required, 2) it delivers unbiased estimation, and 3) the identified model quality can be estimated by inspecting the tracking performance trough experiments. However an important restriction applies to this method, that is only mod-

els with no zeros can be dealt with. In this context several works can be cited, among them [31], [32] and the recent contribute in which the restriction above is removed [33]. It is worth noting that, although linear models are attractive for several reasons, they also have their limitations: most real-life systems show nonlinear dynamic behavior and a linear model can only describe such a system for a small range of input and output values. Therefore, a considerable interest in identification methods for nonlinear systems has risen.
Most work in the area of nonlinear system identification has concentrated on input-output models: a reccomended nice introduction to nonlinear system identifiction was given in [34]. Becuase of the diversity of model structures, the literature on nonlinear systems identification is vast: the reader who wants to delve into it, some pointers to start from are: [35], [36], [37], [38], [39], [40], [41].

This thesis is focused on the identification of linear Single-Input-Single-Output (SISO) CT systems models, based on noisy output measurements; this problem is of great interest in many applications, including radar, sonar, seismics, ocean acustic, communications, control and others. It will be addressed this problem through a kernel-based approach: in particular it will be exploited well established theoretic achievements in the field of algebra of linear integral operators to develop a novel formalism for the synthesis of non-asymptotic parametric and state estimators.

1.2.1 Parameter estimation design techniques

Among the various techniques proposed in literature for CT parametric identification of linear dynamical systems ([42], [14], [16], [43] and the contributed volume [44]), we can recognize two main classes depending on the approach used to overcome the impossibility to measure the derivatives of the input-output signals of the system under concern: *i*) State Variable Filtering (SVF) and *ii*) Integral Methods (IM).
The SVF approach consists in filtering the system's inputs and outputs in order to obtain *prefiltered* time-derivatives in the bandwidth of interest that may be exploited, in place of the unmeasured derivatives of the signals, to estimate the model parameters. Instead, integral methods are related to the proposed methodology and they have quite a long history in the field of continuous-time identification. Among integral techniques, we recall *i*) the Modulating Function (MF) method, which relies on the repeated integration of input-output signals over finite-length intervals to minimize the effect of unknown initial conditions on the estimates; *ii*) the

linear integral filter method, in which the initial conditions must be considered explicitly, by augmenting the dimension of the decision space with the unknown initialization variables; *iii*) the reinitialized partial moments method, that consists in integrating the input-output signals over finite-length time windows, in sampling the integrals, and finally in performing the regression over a discrete time-series, making the overall estimator an inherently hybrid dynamical system.

Typically, in the context of CT identification, asymptotic convergence properties can be proved and several algorithms have been devised to provide good performance in terms of transient behavior of the estimates (see, for example, [45] and the references cited therein). However, in order to achieve estimates' modes of behavior characterized by very fast convergence properties, it is usually necessary to augment the vector of decision variables with the unknown initial conditions of the unmeasured states. The main drawback of this technique is related to numerical issues in estimating the initial hidden states as time goes on.
Within the IM cclas we can also include the Algebraic Estimators recently developed by M.Fliess and co-workers (see [46, 47, 48, 49, 50], the book [51] for the underlying theory and [52, 53, 54, 55] for applications of the method are worth mentioning). While most of the classic CT dynamic estimations techniques rely on model fitting criteria (least-squares, maximum likelyhood), the Algebraic Estimation setting makes use of differential algebra and operational calculus to derive an estimator that provides instantaneous estimates of the system's parameters with a minimal realization. The possibility to obtain non-asymptotic estimates by non-anticipative minimal filtering represents an outstanding peculiar feature of the algebraic approach. Indeed, there is no need to choose initial conditions and first guess estimates, that are conversely critical quantities in classical methods, deeply affecting the transitory behavior of the estimators.
On the other side, the structure of the algebraic estimator is fixed by the synthesis approach, based on the algebraic differentation, in the Laplace's operational domain, of the transformed system's dynamical constraints (in other words, assuming an upper bound on the relative degree of the system, the structural differential constraint is recast in the Laplace's operational domain and then manipulated algebraically to remove the influence of unknown initial conditions on hidden states/output derivatives). By "fixed structure" it is meaning that the estimated quantities can be expressed in a closed-form expression that always involves nested integrals of terms diverging with time. The presence of diverging integrals impacts negatively in the pratical implementation, which always yields to an internally unstable (nonlinear) dynamical system: successive re-initializations are then required for practical implementation, in order to avoid wind-up of the integrators, whose integrands are unbounded functions of time.

On the other side, within the SVF class we include the *Instrumental Variable* (IV) approach to the identification and estimation of trasfer function models: this technique has a rich history in the control and systems literature, with the earliest algorithm of this type dating back to the 1960s ([56], [57], [58], [59], [60]). This techinque allowed to approach the identification of both discrete-time (RIV) and continuous-time (RIVC) Transfer Function (TF) models using the statistically optimal *Refined Instrumental Variable* method ([61], [62], [63], [64], [65], [66]).
The RIV/RIVC algorithms together consitute a unified, time domain family of algorithms that provide statisticaly optimal solutions of both discrete-time and "hybrid" continuous-time TF models of the Box-Jenkins type ([12]). In this regard, they have advantages over alternative algorithms, such as the well knwon *Prediction Error Minimization* (PEM) approach ([14]) used in the $Matlab^{TM}$ System Identification Toolbox, where the general time domain algorithms for *direct* TF model estimation are only available for DT models.

1.2.2 Observers design techniques

The problem of estimating the state of a dynamical system from outputs and inputs (commonly known as "observing the state", hence the name "observer") is an important problem in the theory of systems. Their main function is extracting variables, otherwise unmeasureble, for a vast number of applications including feedback control [67] or system health monitoring [68]. In engineering practice, an observer is used for a number of purposes, such as removing phase lag in feedback, reducing the use of costly sensors [69] and estimating disturbances [70, 71].
Classical observer design techniques for linear continuous-time (CT) systems are characterized by asymptotic convergence of the estimates (see [72, 73]). In many applications, such as fault isolation or change-point detection, the estimates of the hidden states are often required to converge in a neighborhood of the true values within a predetermined finite time, independently from the unknown initial conditions. Several algorithms have been proposed to provide state estimates with finite convergence time. Among others, we recall the moving-horizon observer without a priori information on initial conditions described in [74], the convolutional observer proposed by [75] and the delay-based filters proposed in [76, 77]. The implementation of the aforementioned observers is however difficult or computationally demanding. Indeed, moving-horizon observers need to solve repeated dynamic optimization problems on-line, whose complexity

depends on the system's dimension and on the horizon length. Moreover, huge memory resources are required to construct the delayed signals needed by [76, 77], or to compute numerically the convolution integrals over moving time-windows required by the method described in [75].

A different approach to finite-time state estimation relies on sliding-mode (SM) based update laws (see [78]). Conventional SM observers however can only guarantee the semi-global stability of the estimation-error dynamics, *i.e.*, the convergence can be proven for initial system states contained in a bounded region. Higher-order SM update laws can be shown to achieve global convergence, at the cost of increased implementation complexity. Indeed these observers are expensive from a computational point of view because they require a large (infinite) amount of memory due to the storage of trajectory pieces, and/or the instantaneous solution of convolution integrals over a finite time horizon, and/or by an increased order of the observed dynamics. Notably, the SM methodology achieves finite-time convergence by discontinuous high-gain output injection, so measurement noise may prevent its applicability. An alternative finite-time convergent observer, based on impulsive innovation updates, has been proposed in [79]. This method, originally implemented by dynamic augmentation, has been recently improved by [80] to reduce the dimension of the estimator. The computational complexity of the latter method is reduced with respect to moving-horizon and convolutional approaches, moreover, compared to sliding-mode observes, the stability of the estimation error dynamics can be guaranteed globally and no high-gain output injections are used.

1.3 Publications

The research presented in this thesis has been extensively presented at international conferences and are based on the following publications:

- G.Pin, A.Assalone, M.Lovera and T.Parisini, "Kernel-based Non-Asymptotic Parameter Estimation of Continuous-Time Systems", in Proc. 51st IEEE Conference on Decision and Control, Maui, Hawaii, pp. 2832-2839, 2012.
- G.Pin, M.Lovera, A.Assalone and T.Parisini, "Kernel-Based Non-Asymptotic State Estimation for Linear Continuous-Time Systems", in Proc. 2013 American Control Conference, Washington, pp. 3123-3128, 2013
- G.Pin, A.Assalone, M.Lovera and T.Parisini, "Kernel-based Non-Asymptotic Parameter Estimation of Continuous-time Linear Systems", Submitted in revised form (as a full paper) to the IEEE Transactions on Automatic Control, 2013.

Chapter 2

Methodology

In this Chapter it will be introduced the mathematical background needed in the sequel, fundamental notions and statements of the theory of linear and integral operators will be presented; besides, the Volterra and Fredholm theory is set forth in the Hilbert spaces. It will be analized the fundamental aspects of this field of research, considering the more relevant properties that will be exploited along this work. The notions discussed in this Chapter dont't provide innovative results, therefore this part will be used as a basic instrument, with the purpose to obtain state and parametric estimators with specified properties.

2.1 Linear integral operators

The problems of physics and mechanics relative rarely lead directly to an integral equation; these problems can be described mostly by differential equations. However, many of these dfferential equations can be transformed into integral equations, together with the initial and boundary values. This chapter is aimed at explain with the concept of *integral equations*, focusing our attention in Volterra equations. We will focus in depth the problem of existence for Volterra equations in spaces of continuous/measurable functions. Besides classical Volterra equations involving integral operators, we shall also deal with general Volterra equations that involve causal (or non-anticipative) operators not necessarily of integral type [81] .

2.1.1 Volterra operators

An integral equation is an equation in which the unknown function appears under the integral sign. There is no universal method for solving integral equations. Solution methods and even the existence of a solution depends on the particular form of the integral equation.
An integral equation is called *linear* if linear operations are performed on the unknown function. The *general form* of a linear integral equation is

[82]:

$$g(x)\phi(x) = f(x) + \lambda \int_{a(x)}^{b(x)} K(x,y)\phi(y)dy$$

The unknown function is $\phi(x)$, the function $K(x,y)$ is called the *kernel of the integral function*, $f(x)$ is the so-called *perturbation function* and λ is usually a complex parameter.
Two types of integral equation are of special importance. If the limits of the integral are independent, *i.e.*, $a(x) = a$ and $b(x) = n$, we call it a *Fredholm integral equation*.
If $a(x) = a$ and $b(x) = x$, we call it a **Volterra integral equation**. Moreover, if the unknown function $\phi(x)$ appears only under the integral sign, *i.e.*, $g(x) = 0$ holds, we have an integral equation of the *first kind* and takes the form:

$$0 = f(x) + \lambda \int_a^b K(x,y)\phi(y)dy, \qquad 0 = f(x) + \lambda \int_a^x K(x,y)\phi(y)dy$$

Instead, if $g(x) = 1$ the equation is called *of the second kind*, as:

$$\phi(x) = f(x)+\lambda \int_a^b K(x,y)\phi(y)dy, \qquad \phi(x) = f(x)+\lambda \int_a^x K(x,y)\phi(y)dy$$

In the following we will focus our attention in the second type of integral equations that will be used in our methodology.

A Volterra integral equation has the form:

$$\phi(x) = f(x) + \int_a^x K(x,y)\phi(y)dy \tag{2.1}$$

The solution function $\phi(x)$ with the independent variable x from the closed interval $I = [a,b]$ or from the semi-open interval $I = [a,\infty]$ is required; the following theorem provides the existence of a unique solution of the integral equation [82].

Theorem 2.1.1 *If the functions $f(x)$ for $x \in I$ and $y \in [a,x]$ are continuous, then there exists a unique solution $\phi(x)$ of the integral equation such that it is continuous for $x \in I$*

The Volterra integral equation of the first kind can be transformed into an equation of the second kind. Hence, theorems about existence and uniqueness of the solution are valid with some modifications.

Remark 2.1.1 (Transformation by differentiation) *Assuming that $\phi(x)$, $K(x,y)$ are continuous functions, we can tranform the integral equation of the first kind*

$$f(x) = \int_a^x K(x,y)\phi(y)dy \tag{2.2}$$

into the form

$$f^{'}(x) = K(x,x)\phi(x) + \int_a^x \frac{\partial}{\partial x}K(x,y)\phi(y)dy$$

by differentiation with respect to x. If $K(x,x) \neq 0 \ \forall \ x \in I$, then dividing the equation by $K(x,x)$ we get an integral equation of the second kind.

Remark 2.1.2 (Transformation by partial integration) *Assuming that $\phi(x)$, $K(x,y)$ are continuous, we can evaluate the integral* (2.2) *by partial integration. Substituting*

$$\int_a^x \phi(y)dy = \psi(x)$$

gives

$$\begin{aligned} f(x) &= \Big[K(x,y)\psi(y)\Big]_{y=a}^{y=x} - \int_a^x \Big(\frac{\partial}{\partial y}K(x,y)\Big)\psi(y)dy \\ &= K(x,x)\psi(x) - \int_a^x \Big(\frac{\partial}{\partial y}K(x,y)\Big)\psi(y)dy \end{aligned}$$

If $K(x,x) \neq 0$ for $x \in I$, then dividing by $K(x,x)$ we have an integral equation of the second kind:

$$\psi(x) = \frac{f(x)}{K(x,x)} + \frac{1}{K(x,x)}\int_a^x \Big(\frac{\partial}{\partial y}K(x,y)\Big)\psi(y)dy$$

Differentiating the solution $\psi(x)$ we get the solution $\phi(x)$ of (2.2).

Most ordinary differential equations can be expressed as integral equation, but the reverse is not true [83]. Given an $n-$th order differential equation

$$x^{(n)}(t) = f(t, x, x^{'}, \cdots, x^{(n-1)})$$

it is possible to express this differential equation as a system of n first order equations and then formally integrated. For example if $x^{''} = f(t,x,x^{'})$, then it is possible to assign $x = x_1$ and $x^{'} = x_1^{'} = x_2$, such that $x^{''} = x_2^{'} = f(t, x_1, x_2)$, in vectorial notation:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ f(t, x_1, x_2) \end{bmatrix}$$

Therefore, in general, if $\mathbf{x} \in R^n$ then:

$$\mathbf{x}^{'} = \mathbf{G}(t, \mathbf{x}), \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

is a system of n first order equations with initial conditions (called an *initial − value problem*), written as

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{G}(s, \mathbf{x}(s))ds$$

is a system of n integral equations. Thus, it is trivial to express such differential equations as integral equations, it is mainly a matter of renaming variables.
As we already mentioned, the more general case of integral operator is called Fredholm integral operator, and its kernel is called *Fredholm extension* of the Volterra kernel $K(t, \tau)$. The Fredholm kernel extension can be exploited to specialize to Volterra operators the properties and the results conceived within the Fredholm operator theory. In particular, we will use the Fredholm kernel extension to characterize the kernel of composed Volterra operators (see the Appendix).

A *signal* is defined as a generic function of time $u(t) : t \mapsto u_t$, $u_t \in \mathbb{R}$, such that $u(\cdot) \in \mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0})$. Furthermore, given two scalars $a, b \in \mathbb{R}_{\geq 0}$, with $a < b$, let us denote by $u_{[a,b]}(\cdot)$ and $u_{(a,b]}(\cdot)$ the *restriction* of a signal $u(\cdot)$ to the closed interval $[a, b]$ and to the left-open interval $(a, b]$, respectively. Indeed, we have the following:

Definition 2.1.1 (Weak (generalized) Derivative) *Let $u(\cdot) \in \mathcal{L}^1_{loc}(\mathbb{R}_{\geq 0})$. We say that $u^{(1)}(\cdot)$ is a weak derivative of $u(\cdot)$ if*

$$\int_0^t u(\tau)\left(\frac{d}{d\tau}\phi(\tau)\right)d\tau = -\int_0^t u^{(1)}(\tau)\phi(\tau)d\tau, \;\; \forall t \in \mathbb{R}_{\geq 0}$$

for all $\phi \in C^\infty$, with $\phi(0) = \phi(t) = 0$. □

We remark that $u^{(1)}(.)$ is unique up to a set of Lebesgue measure zero, *i.e.*, it is defined almost everywhere. If $u(.)$ is differentiable in the conventional sense, then its weak derivative is identical to its conventional derivative. Classical rules for the derivation of sum or products of functions also hold for the weak derivative. In analogy with the conventional derivative, we will denote the $i-$th derivative (if exists) as:

$$u^{(i)}(\tau) = \frac{d^i}{d\tau^i}u(\tau)$$

with $i \in \mathbb{Z}_{\geq 0}$.
The notion of generalized partial derivative may be extended to functions of many variables: given a kernel function $K(.,.) \in \mathcal{HS}$, we will denote the weak derivative with respect to the second argument (if exists) as:

$$K^{(i)}(t,\tau) = \frac{\partial}{\partial \tau^i} K(t,\tau)$$

Let us also introduce the integral function:

$$u^{[1]}(\tau) = \int_0^\tau u(\sigma) d\sigma$$

and the $i-$th integral function

$$u^{[i]}(\tau) = \int_0^\tau \int_0^{\sigma_{i-1}} \cdots \int_0^{\sigma_3} \int_0^{\sigma_2} K(\sigma_1) d\sigma_1 d\sigma_2 \cdots d\sigma_{i-1} d\sigma_i$$

Moreover, for a $\mathcal{HS}$ kernel, let

$$K^{[i]}(t,\tau) = \int_0^\tau \int_0^{\sigma_{i-1}} \cdots \int_0^{\sigma_3} \int_0^{\sigma_2} K(t,\sigma_1) d\sigma_1 d\sigma_2 \cdots d\sigma_{i-1} d\sigma_i$$

where we can note that $K^{[0]}(t,\tau) = K^{(0)}(t,\tau) = K(t,\tau)$, $\forall t \in \mathbb{R}_{\geq 0}$, $\forall \tau \in \mathbb{R}_{\geq 0}$
Finally, the notion of BIBO stability for an integral operator is introduced.

Definition 2.1.2 (BIBO Stability) *A bounded linear operator T, $T \in \mathcal{B}(\mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0}),\, \mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0}))$ is said BIBO-stable if:*

$$\Big|[Tx](t)\Big| < \infty, \forall t \in \mathbb{R}_{\geq 0}, \forall x(\cdot) \in \mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0}) : \big\{ \big|x(\tau)\big| < \infty, \forall \tau \in \mathbb{R}_{\geq 0} \big\}.$$

□

In the case of a Volterra operator V_K, BIBO stability is equivalent to the following property of the kernel:

$$\sup_{t \in \mathbb{R}_{>0}} \left\{ \int_0^t |K(t,\tau)| d\tau \right\} < \infty. \tag{2.3}$$

Condition (2.3) will be used in the sequel to assess the stability of the operators in our setting. A kernel fulfilling (2.3) will be called a BIBO *stable kernel.* In this respect, it is worth noting that BIBO stability *per se* is not sufficient to establish the existence of a finite-dimensional state-space realization for an operator, that is, its implementability. The order of the realization can be determined only when an analytical expression for the kernel is available.

2.1.2 Operator Theory and Hilbert Spaces

Intuitively, a system is a black box whose input and output are function of time (or vectors of such function): as such, a natural model for a system is an operator defined on a function space. This observation has the effect that system theory is a subset of operator theory ([84]). The difficulties lie in the fact that the operators encountered in systems theory are defined on spaces of time functions and, as such, must satisfy a physical realizability (**causality**) condition to the effect that the operator cannot predict the future.
In an effort to alleviate these and similar problems, encountered for instance in the design of regulators, passive filters and stochastic systems, the theory of operators defined on a Hilbert resolution space was developed in the mid-1960s. In essence, a Hilbert **resolution space** is simply a Hilbert space to which a time structure has been axiomatically adjoined, thereby, allowing one to define such concepts as causality, stability, memory and passivity in an operator theoretic settings([6],[84]).
A real linear space R with the *inner product* satisfying

$$\begin{aligned} (x, y) &= (y, x) \quad \forall x, y \in \mathbb{R}, \\ (x_1 + x_2, y) &= (x_1, y) + (x_2, y), \quad \forall x_1, x_2 \in \mathbb{R}, \\ (\lambda x, y) &= \lambda(x, y), \quad \forall x, y \in \mathbb{R}, \\ (x, x) &\geq 0, \quad \forall x \in \mathbb{R}, \quad (x, x) = 0 \quad \iff \quad x = 0. \end{aligned} \tag{2.4}$$

is called the *Euclidian space*.
A Hilbert space H is a complete infinite-dimensional Euclidian space and the operator A^* *adjoint* to an operator A is defined by

$$(Ax, y) = (x, A^*y), \qquad \forall x, y \in H$$

In the context of Integral equations, Hilbert space fullfill an important role. Let's consider the integral equation (2.1) and let's assume that $K(x, y)$ is a Hilbert-Schmidt kernel, *i.e.*, a square-integrable function in the square $Pi = \{(x, y) : a \leq x \leq b, a \leq y \leq b\}$, so that:

$$\int_a^b \int_a^b |K(x, y)|^2 dx dy \leq \infty \tag{2.5}$$

and $f(x) \in L_2[a, b]$, *i.e.*

$$\int_a^b |f(x)|^2 dx \leq \infty$$

Defining a linear Fredholm integral operator corrisponding to (2.1), we have:

$$A\phi(x) = \int_a^b K(x, y)\phi(y) dy \tag{2.6}$$

If $K(x, y)$ is a Hilbert-Schimdt kernel, then the operator (2.6) will be called a *Hilbert-Schimdt operator*.
Rewriting now (2.1) as a linear operator equation:

$$\phi = A\phi(x) + f, \qquad f, \phi \in L_2[a, b] \tag{2.7}$$

we can derive the result of the following theorem (see [85] and [86]).

Theorem 2.1.2 *Equality* (2.7) *and condition* (2.6) *define a* ***complitely continuous linear operator*** *in the space* $L_2[a, b]$

2.1.3 Non-anticipativity and non-asymptotic Volterra operator

The notions of causality and non-anticipativity play a key role in characterizing the implementability (existence of a stable finite-dimensional state-space realization for the integral operators) of the proposed methodology [3], [4]. In non-formal terms, an operator
$T \in \mathcal{B}(\mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0}), \mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0}))$ is said to be causal (non-anticipative) if at any time $t > 0$ (respectively, $t \geq 0$) the image of a signal $x(\cdot)$ at time t, $[Tx]\,(t)$, depends only on the restriction $x_{[0,t)}(\cdot)$ (respectively, $x_{[0,t]}(\cdot)$). Being the Volterra operator inherently non-anticipative, the signal $[V_K x]\,(t)$, for $t \geq 0$, can be obtained as the output of a dynamic system described by the following scalar integro-differential equation:

$$\begin{cases} \xi^{(1)}(t) = \begin{cases} K(t,t)x(t) + \displaystyle\int_0^t \left(\frac{\partial}{\partial t}K(t,\tau)\right) x(\tau)d\tau, & t \in \mathbb{R}_{>0}; \\ 0, & t = 0; \end{cases} \\ \xi(0) = \xi_0 = \displaystyle\int_0^0 K(0,\tau)x(\tau)d\tau; \end{cases}$$
$$[V_K x]\,(t) = \xi(t), \quad \forall t \in \mathbb{R}_{>0}, \tag{2.8}$$

where $\xi^{(1)}(t) = \frac{d}{dt}[V_K x](t)$ has been obtained by applying the Leibnitz rule in deriving the integral.

Now, we will introduce some useful results dealing with the application of Volterra operators to the derivatives of a signal.

Lemma 2.1.1 (Volterra Integral of a function derivative) *For a given $i \geq 0$, consider a signal $x(\cdot) \in \mathcal{L}^2(\mathbb{R}_{\geq 0})$ that admits a i-th weak derivative in $\mathbb{R}_{\geq 0}$ and a kernel function $K(\cdot, \cdot) \in \mathcal{HS}$ that admits the i-th derivative (in the conventional sense) with respect to the second argument, $\forall t \in \mathbb{R}_{\geq 0}$. Then, it holds that:*

$$
\begin{aligned}
\left[V_K x^{(i)}\right](t) = &\sum_{j=0}^{i-1}(-1)^{i-j-1}x^{(j)}(t)K^{(i-j-1)}(t,t) \\
&+\sum_{j=0}^{i-1}(-1)^{i-j}x^{(j)}(0)K^{(i-j-1)}(t,0) \\
&+(-1)^i\left[V_{K^{(i)}}\, x\right](t), \;\; \forall t \in \mathbb{R}_{\geq 0},
\end{aligned} \tag{2.9}
$$

that is, the function $\left[V_{K_h} x^{(i)}\right](\cdot)$ *is non-anticipative with respect to the lower-order derivatives* $x(\cdot), \ldots, x^{(i-1)}(\cdot)$. □

Proof 1 *The following result can be obtained by integrating by parts:*

$$
\begin{aligned}
\left[V_K x^{(i)}\right](t) &= \int_0^t K(t,\tau)x^{(i)}(\tau)d\tau \\
&= x^{(i-1)}(t)K(t,t) - x^{(i-1)}(0)K(t,0) \\
&\quad -\int_0^t K^{(1)}(t,\tau)x^{(i-1)}(\tau)d\tau.
\end{aligned} \tag{2.10}
$$

The integral operator on the right-hand side of (2.10) can be further split by parts:

$$
\begin{aligned}
&-\int_0^t K^{(1)}(t,\tau)x^{(i-1)}(\tau)d\tau = \\
&= -x^{(i-2)}(t)K^{(1)}(t,t) + x^{(i-2)}(0)K^{(1)}(t,0) \\
&\quad +\int_0^t K^{(2)}(t,\tau)x^{(i-2)}(\tau)d\tau.
\end{aligned}
$$

Proceeding by induction we obtain the expression

$$
\begin{aligned}
\int_0^t K(t,\tau)x^{(i)}(\tau)d\tau = &\sum_{j=1}^{i}(-1)^{j+1}x^{(i-j)}(t)K^{(j-1)}(t,t) \\
&+\sum_{j=1}^{i}(-1)^{j}x^{(i-j)}(0)K^{(j-1)}(t,0) \\
&+(-1)^i\int_0^t K^{(i)}(t,\tau)x(\tau)d\tau
\end{aligned} \tag{2.11}
$$

that is, the function obtained by applying the Volterra operator to the i-th derivative is non-anticipative with respect to lower-order derivatives.

Finally, by rearranging the indexing of the summation in the right hand side of (2.11), the statement of the lemma trivially follows.

By exploiting the identity (2.9), we now characterize the kernels for which the transformed signal $\left[V_K x^{(i)}\right](\cdot)$ is independent from the initial values of the derivatives $x(0), x^{(1)}(0), \ldots, x^{(i-1)}(0)$. The following definition characterizes the kernels yielding non-asymptotic Volterra operators.

Definition 2.1.3 (i-th Order Non-Asymptotic Kernel) *Consider a kernel $K(\cdot,\cdot)$ satisfying the assumptions posed in the statement of Lemma 2.1.1; if for a given $i \geq 1$, the kernel verifies the supplementary condition*

$$K^{(j)}(t,0) = 0, \quad \forall t \in \mathbb{R}_{\geq 0}, \, \forall j \in \{0,\ldots,i-1\},$$

then, it is called an i-th order non-asymptotic kernel. □

Lemma 2.1.2 (Non-asymptoticity Implication)
If a kernel $K(\cdot,\cdot)$, is at least i-th order non-asymptotic, then the image function of $x^{(i)}(\cdot)$ at time t, $\left[V_K x^{(i)}\right](t)$, depends only on the instantaneous values of the lower-order derivatives $(x(t), x^{(1)}(t), \ldots, x^{(i-1)}(t))$ and on the restriction $x_{(0,t]}(\cdot)$, but not on the initial states $x(0), x^{(1)}(0), \ldots, x^{(i-1)}(0)$. □

The proof of Lemma 2.1.2 follows immediately from Lemma 2.1.1 and is therefore omitted.

Up to now, we have characterized a candidate class of kernels which allows to remove the influence of the unknown initial derivatives from the transformed signal $\left[V_K x^{(i)}\right](t)$. However, beyond depending on the current value $x(t)$ and its past time-behaviour, such a signal depends also on the unmeasurable instantaneous values of the lower-order derivatives $x^{(j)}(t)$, with $j \in \{1,\ldots,n-1\}$. To address this issue, we need to introduce the notion of composed (or nested) Volterra operators and to discuss some relevant properties.

Let us denote by $\left[V_{K_N \bullet \cdots \bullet K_1} x^{(i)}\right](\cdot)$, the image function obtained by composing N Volterra integral operators to $x^{(i)}(\cdot)$:

$$\left[V_{K_N \bullet \cdots \bullet K_1} x^{(i)}\right] = \left[V_{K_N} \cdots \left[V_{K_2}\left[V_{K_1} x^{(i)}\right]\right]\right].$$

In view of the composition property of Volterra operators (see (8.1) and (8.2) in the Appendix), it holds that the composed operator is in turn a Volterra operator with kernel $K_N \bullet K_{N-1} \bullet \cdots \bullet K_i \bullet \cdots \bullet K_2 \bullet K_1$, where $\cdot \bullet \cdot$ denotes the kernel-composition integral (see (8.2) in the Appendix).

Theorem 2.1.3 (Non-asymptotic Derivative Image) *Let $x^{(i)}(\cdot)$ be the i-th derivative of the signal $x(\cdot)$ and let $N \geq i$ be an arbitrary integer. Given N kernel functions $K_1(\cdot,\cdot), \ldots, K_N(\cdot,\cdot)$, such that K_1 is d-th order non-asymptotic, with $d \geq i-1$ and $K_j \in \mathcal{HS}$, $\forall j \in \{1,\ldots,N\}$, consider the composed operator V_{P_N}, with kernel*

$$P_N \triangleq K_N \bullet \cdots \bullet K_2 \bullet K_1.$$

The image of $x^{(i)}(\cdot)$ through V_{P_N}, $[V_{P_N}x^{(i)}](\cdot)$, can be obtained as the image of the restriction $x_{[0,t]}(\cdot)$ through a non-anticipative operator. Indeed, there exists an operator $V_{R_{N,i}}$, induced by the kernel

$$R_{N,i} \triangleq -K_N \bullet K_{N-1} \bullet \cdots \bullet K_{i+1} \bullet T_i,$$

with $T_i(\cdot,\cdot)$ defined recursively by

$$\begin{cases} T_j(t,\tau) & \triangleq -\Big(K_j \bullet T_{j-1}^{(1)}\Big)(t,\tau) + K_j(t,\tau)T_{j-1}(\tau,\tau), \\ \quad T_1 & \triangleq K_1, \qquad\qquad \forall j \in \{2,\ldots,i\},\ \forall (t,\tau) \in \mathbb{R}^2, \end{cases} \tag{2.12}$$

such that

$$\big[V_{P_N}x^{(i)}\big](t) = R_{N,i}(t,t)x(t) - R_{N,i}(t,0)x(0) - \big[V_{R_{N,i}^{(1)}}\,x\big](t). \tag{2.13}$$

$\square$

Proof 2 *First, by integrating by parts, the innermost operator can be decomposed as*

$$\big[V_{P_N}x^{(i)}\big](t) = \Big[V_{K_N}\cdots\Big[V_{K_2}\Big(\tilde{x}_1 - \big[V_{K_1^{(1)}}\,x^{(i-1)}\big]\Big)\Big]\Big](t) \tag{2.14}$$

where $\tilde{x}_1(t) = K_1(t,t)x^{(i-1)}(t) - K_1(t,0)x^{(i-1)}(0)$. Consider now the composed kernel

$$T_2(t,\tau) = -(K_2 \bullet K_1^{(1)})(t,\tau) + K_2(t,\tau)K_1(\tau,\tau),$$

obtained by (2.12) (recall that $K_2 \bullet K_1^{(1)} = \int_\tau^t K_2(t,\sigma)K_1^{(1)}(\sigma,\tau)d\sigma$, see also the Appendix). By the non-asymptoticity property of K_1: $K_1(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}$, and in view of (2.12) and (2.14) we obtain

$$[V_{P_N}x^{(i)}] = \Big[V_{K_N}\cdots\Big[V_{K_3}\big[V_{T_2}(-x^{(i-1)})\big]\Big]\Big](t). \tag{2.15}$$

Integrating by parts, the innermost operator in (2.15) can be decomposed as

$$[V_{P_N}x^{(i)}] = \Big[V_{K_N}\cdots\Big[V_{K_3}\big(\tilde{x}_2 + [V_{T_2^{(1)}}x^{(i-2)}]\big)\Big]\Big](t),$$

where $\tilde{x}_2(t) \triangleq -T_2(t,t)x^{(i-2)}(t) + T_2(t,0)x^{(i-2)}(0)$. Since the kernel $K_1(\cdot,\cdot)$ subsumes the i-th order non-asymptoticity condition, then, by (2.12), $T_2(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}$, i.e., *also $T_2(\cdot,\cdot)$ is non-asymptotic. We can then write*

$$\big[V_{P_N}x^{(i)}\big](t) = \Big[V_{K_N}\cdots\Big[V_{K_4}\big[V_{T_3}(-x^{(i-2)})\big]\Big]\Big](t).$$

Integrating again by parts, the innermost operator can be decomposed as

$$\big[V_{P_N}x^{(i)}\big](t) = \Big[V_{K_N}\cdots\Big[V_{K_4}\big(\tilde{x}_3 + \big[V_{T_3^{(1)}}x^{(i-3)}\big]\big)\Big]\Big](t)$$

where $\tilde{x}_3(t) \triangleq -T_3(t,t)x^{(i-3)}(t) + T_3(t,0)x^{(i-3)}(0)$. *Due to the fact that the kernel* $K_1(\cdot,\cdot)$ *is at least* $i-1$*-th order non-asymptotic, then* $T_j(t,0) = 0, \forall j \in 1,\ldots,i-1$. *By iterating this line of reasoning we get*

$$\begin{aligned}\left[V_{P_N} x^{(i)}\right](t) &= \left[V_{K_N} \cdots \left[V_{K_{i+1}}\left[V_{T_i}(-x^{(1)})\right]\right]\right](t)\\ &= \left[V_{R_{N,i}} x^{(1)}\right](t)\\ &= R_{N,i}(t,t)x(t) - R_{N,i}(t,0)x(0) - \left[V_{R^{(1)}_{N,i}} x\right](t).\end{aligned}$$

The following Remark sheds some light on significant implications of Theorem 2.1.3.

Remark 2.1.3 (Implications) *In Theorem 2.1.3, the existence of a composed Volterra integral operator has been shown, namely* $V_{P_N} = V_{K_N \bullet \cdots \bullet K_1}$, *that, fed by the* $i-th$ *derivative* $x^{(i)}(\cdot)$ *of a signal, produces an image signal, say* $\left[V_{P_N} x^{(i)}\right](\cdot)$, *which, in turn, can be expressed, in the most general case, in terms of the sole restriction* $x_{[0,t]}(\cdot)$ *(or* $x_{(0,t)}(\cdot)$ *under slightly stronger assumptions) and that, in any case, does not depend on the initial conditions of the hidden derivatives. Assume now that* $x^{(i)}(\cdot)$ *is not measurable while* $x(\cdot)$ *is available; then, thanks to (2.13), the signal* $\left[V_{P_N} x^{(i)}\right](\cdot)$ *can be obtained by applying a non-anticipative operator (see (2.13)) to the measurable signal* $x(\cdot)$.

2.2 Concluding remarks

In this chapter has been illustrated the fundamental theoretical background of the proposed methodology. Volterra integrals and its properties have been introduced, providing the basis for the further discussion. In the following, the proposed approach will be formalized, analytic details and results will be derived for both parametric and state estimation.

Part I

Kernel-based Parameter Estimation

Chapter 3

Non-asymptotic kernels for parameter estimation

This Chapter aims at deriving some analytic instances of non-asymptotic kernels. In particular, we will show that the Bounded-Input-Bounded-Output (BIBO) stability of the Volterra operators induced by the kernels reflects on the stability properties of the estimator. It will be considered the unstable and stable realizations given by a class of univariate kernel functions (U-NK), implementable by time-weighted integrals of the input and output signals. It will be shown that in the U-NK framework it is possible to enforce the internal stability of the non-asymptotic estimator by a suitable choice of the kernels. However, the internal stability of U-NK-based estimators can only be enforced by damping the kernels asymptotically, therefore internal stability comes at the price of a practical freeze of the estimator as time proceeds.
To avoid the problem of estimator-freeze, it will be presented a relevant contribution of the work, consisting in the analytical expression of a Bivariate Causal Non-Asymptotic Kernel (BC-NK). The operators induced by the proposed BC-NK yield an estimator that admits a finite-dimensional time-varying linear state-space realization. Moreover, the dynamics of BC-NK-based estimator asymptotically tend to a time-invariant (stable) linear dynamical system, in which input and output injections are never suppressed.

3.1 Preliminares

Let's consider a general SISO (Single Input Single Output) Continuous-time system $\mathcal{S}_{u\to y}$, whence y and u represent respectively the output and the input of the system:

$$\begin{cases} y^{(n)}(t) = \sum_{i=0}^{n-1} a_i y^{(i)}(t) + \sum_{k=0}^{m-1} b_k u^{(k)}(t) \forall t \in \mathbb{R}_{\geq 0}; \\ y^{(i)}(0) = y_0^{(i)},\ i \in \{0,\ldots,n-1\}; \\ u^{(k)}(0) = u_0^{(k)},\ k \in \{0,\ldots,m-1\} \end{cases} \tag{3.1}$$

with $m \in \mathbb{Z}_{>0}$, $n \in \mathbb{Z}_{>0}$, $m \leq n$ and $p \in \mathbb{Z}_{\geq 0}$. The values of the constant parameters $a_i \in \mathbb{R}$, $i \in \{0, \dots, n-1\}$, $b_k \in \mathbb{R}$, $k \in \{0, \dots, m-1\}$ are unknown. The only measurable signals are $y(t)$ and $u(t)$, while their time-derivatives are not assumed to be available. Our objective consists in estimating the system's parameters a_i and b_k by suitably processing the input and output signals $u(t)$ and $y(t)$.

Let's now focus on the structural constraint (3.1) which relates the unknown parameters with the time-derivatives of the signals $u(\cdot)$ and $y(\cdot)$. In the sequel, the results presented in the previous section will be exploited to overcome the unavailability of signal derivatives (hidden internal states of the system) in (3.1), thus obtaining non-asymptotic estimates of the unknown parameters by means of causal filtering.

First, in order to get rid of the structured perturbation term, let us take the p-th generalized derivative of both sides of the structural constraint, obtaining:

$$y^{(n+p)}(t) = \sum_{i=0}^{n-1} a_i y^{(i+p)}(t) + \sum_{k=0}^{m-1} b_k u^{(k+p)}(t). \tag{3.2}$$

Moreover, after choosing an integer $N \geq n + p$, let us apply the Volterra operator $V_{P_N} = V_{K_N \bullet \cdots \bullet K_1}$ (with kernels taken as in Theorem 2.1.3) to both sides of (3.2):

$$[V_{P_N} y^{(n+p)}](\cdot) = \sum_{i=0}^{n-1} a_i [V_{P_N} y^{(i+p)}](\cdot) + \sum_{k=0}^{m-1} b_k [V_{P_N} u^{(k+p)}](\cdot). \tag{3.3}$$

In view of (2.13), we can rewrite (3.3) as

$$r_{y,n+p}(t) = \sum_{i=0}^{n-1} a_i r_{y,i+p}(t) + \sum_{k=0}^{m-1} b_k r_{u,k+p}(t), \; \forall t \in \mathbb{R}_{\geq 0}, \tag{3.4}$$

where the auxiliary signals in (3.4) can be obtained as the image of measurable signals $y(\cdot)$ and $u(\cdot)$ through non-anticipative operators:

$$\begin{aligned} r_{y,j}(t) &= R_{N,j}(t,t)y(t) - R_{N,j}(t,0)y(0) - [V_{R^{(1)}_{N,j}} y](t), \\ r_{u,j}(t) &= R_{N,j}(t,t)u(t) - R_{N,j}(t,0)u(0) - [V_{R^{(1)}_{N,j}} u](t), \end{aligned} \tag{3.5}$$

with $j \in \{p, \dots, n+p\}$ and $j \in \{p, \dots, m+p-1\}$, respectively. Finally, by introducing the true parameter vector

$$\boldsymbol{\theta}^* \triangleq [a_0, \dots, a_{n-1}, b_0, \dots, b_{m-1}]^\top,$$

and the vector of auxiliary signals

$$\boldsymbol{z}(t) \triangleq [r_{y,p}(t), \dots, r_{y,p+n-1}(t), r_{u,p}(t), \dots, r_{u,p+m-1}(t)]^\top,$$

equation (3.4) can be rewritten in compact notation as

$$\boldsymbol{z}^\top(t)\boldsymbol{\theta}^* = r_{y,n+p}(t), \;\; t \in \mathbb{R}_{\geq 0}. \tag{3.6}$$

Now, assuming that all the operators in our formulation admit a stable realization, we need to collect a suitable number of equations like (3.4) in order to form a well-posed algebraic system, to be solved in the unknown parameters. Several approaches can be used to obtain the needed set of constraints, among which we will discuss the following *augmentation methods*: *A*) time sampling, *B*) successive integration, and *C*) covariance filtering.

Time Sampling

This method consists in collecting $N_s \geq n+m$ samples of $r_{y,n+p}(\cdot)$ and $\boldsymbol{z}(\cdot)$ (see (3.6)) at the time instants $t-T_1, t-T_2, \ldots, t-T_{N_s}$ with $0 \leq T_1 < T_2 < \ldots < T_{N_s} \leq t$, and forming, at time t, the algebraic system

$$\underbrace{\begin{bmatrix} r_{y,n+p}(t-T_1) \\ \vdots \\ r_{y,n+p}(t-T_{N_s}) \end{bmatrix}}_{\boldsymbol{r}(t)} = \underbrace{\begin{bmatrix} \boldsymbol{z}^\top(t-T_1) \\ \vdots \\ \boldsymbol{z}^\top(t-T_{N_s}) \end{bmatrix}}_{\boldsymbol{Z}(t)} \boldsymbol{\theta}^*, \tag{3.7}$$

which is known to be satisfied at any time instant by the parameter vector $\boldsymbol{\theta}^*$. An estimate $\hat{\boldsymbol{\theta}}(t)$ of the parameter vector $\boldsymbol{\theta}^*$ can be obtained by minimizing the quadratic fitting criterion

$$\hat{\boldsymbol{\theta}}(t) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{m+n}} ||\boldsymbol{Z}(t)\boldsymbol{\theta} - \boldsymbol{r}(t)||_2, \tag{3.8}$$

that admits the well-known solution $\hat{\boldsymbol{\theta}}(t) = (\boldsymbol{Z}(t))^\dagger \boldsymbol{r}(t)$, where $(\cdot)^\dagger$ denotes the Moore-Penrose matrix pseudoinverse. In the case $N_s = m+n$ and $\boldsymbol{Z}(t)$ is invertible, then the solution can be simply written as $\hat{\boldsymbol{\theta}} = (\boldsymbol{Z}(t))^{-1} \boldsymbol{r}(t)$. A sufficient condition for the existence of a unique solution to the estimation problem (3.8) at a time instant $t \in \mathbb{R}_{>0}$ is that the following *excitation condition* holds:

$$\boldsymbol{Z}^\top(t)\boldsymbol{Z}(t) > 0. \tag{3.9}$$

Therefore, it is possible to establish the existence of a solution to the estimation problem at time t by checking the conditioning of a matrix. Remarkably, in order to ensure the well-posedness of the problem, the time-sampling method requires that a sufficient number of samples has been collected.

Successive Integration

This methods consists in computing supplementary auxiliary signals by the cascaded application of Volterra operators, in order to augment (3.6) with $N_s - 1$ transformed equations that are verified at any time t by the parameter vector $\boldsymbol{\theta}^*$:

$$\underbrace{\begin{bmatrix} r_{y,n+p}(t) \\ [V_G\, r_{y,n+p}]\,(t) \\ [V_{G\bullet G}\, r_{y,n+p}]\,(t) \\ \vdots \\ [V_{G\bullet\cdots\bullet G}\, r_{y,n+p}]\,(t) \end{bmatrix}}_{\boldsymbol{r}(t)} = \underbrace{\begin{bmatrix} \boldsymbol{z}^\top(t) \\ \left[V_G\, \boldsymbol{z}^\top\right](t) \\ \left[V_{G\bullet G}\, \boldsymbol{z}^\top\right](t) \\ \vdots \\ \left[V_{G\bullet\cdots\bullet G}\, \boldsymbol{z}^\top\right](t) \end{bmatrix}}_{\boldsymbol{Z}(t)} \boldsymbol{\theta}^*, \tag{3.10}$$

where the kernel $G(\cdot,\cdot)$ can be chosen arbitrarily. The excitation condition posed for the sampling method, (3.9), must be assumed also in this case to guarantee the existence of a unique solution $\hat{\boldsymbol{\theta}}(t)$ to the estimation problem (3.8). Any input-output signal pair $(u(\cdot), y(\cdot))$ such that $\boldsymbol{Z}^\top(t)\boldsymbol{Z}(t) > 0$, will be addressed to as *sufficiently informative at time time t* in the sequel[1]. In nominal conditions, if the measured signals are sufficiently informative at time t, then the estimate is exact, independently from the initial conditions of the hidden states of the system (*i.e.*, non-asymptotically).

Remark 3.1.1 (Combined Sampling and Integration) *It is interesting to point out that time sampling and successive integration can be also used in a combined way, as proposed in other approaches to CT model identification, see,* e.g.*, [87, 88].*

Covariance Filtering

This method consists in forming the instantaneous covariance equation by left-multiplying (3.6) by $\boldsymbol{z}(t)$:

$$\boldsymbol{R}(t)\boldsymbol{\theta}^* = \boldsymbol{S}(t), \;\; \forall t \in \mathbb{R}_{\geq 0}, \tag{3.11}$$

where $\boldsymbol{R}(\cdot) \in [\mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0})]^{(n+m)\times(n+m)}$ and $\boldsymbol{S}(\cdot) \in [\mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0})]^{n+m}$ are the so-called auto-covariance and cross-covariance matrices defined as

$$\begin{array}{rl} \boldsymbol{R}(t) & \triangleq \boldsymbol{z}(t)\boldsymbol{z}^\top(t) \\ \boldsymbol{S}(t) & \triangleq \boldsymbol{z}(t) r_{y,n+p}(t) \end{array}, \quad \forall t \in \mathbb{R}_{\geq 0}.$$

The instantaneous auto-covariance matrix $\boldsymbol{R}(t)$ is rank-one and therefore it can never be inverted when $n+m \geq 2$. In order to get an invertible system

[1]We remark that the estimated vector $\hat{\boldsymbol{\theta}}(t)$ can be time-varying, because the signals processed by the operators may become sufficiently informative only after a period of time $T \in \mathbb{R}_{>0}$. That is, $\hat{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}^*, \forall t \geq T$, while for $t < T$ the estimate remains undetermined.

for some suitable input-output signal pair, we apply a further Volterra operator V_G, with an arbitrary non-negative kernel $G(\cdot,\cdot) \in \mathcal{HS} : G(t,\tau) \geq 0,\ \forall (t,\tau) \in \mathbb{R}^2_{\geq 0}$, on both sides of (3.11), obtaining the filtered covariance equation:

$$[V_G\, \boldsymbol{R}](t)\boldsymbol{\theta}^* = [V_G\, \boldsymbol{S}](t),\ \ \forall t \in \mathbb{R}_{\geq 0}, \tag{3.12}$$

where the operator V_G has to be applied element-wise on $\boldsymbol{R}(\cdot)$ and $\boldsymbol{S}(\cdot)$. Letting

$$\mathbf{Z}(t) = [V_G \mathbf{R}](t)\,, \quad \mathbf{r}(t) = [V_G \mathbf{S}](t)$$

an optimal estimation problem, analogous to (3.8), can be solved for the estimated parameter vector $\hat{\boldsymbol{\theta}}(t)$. In this case, the invertibility of the filtered auto-covariance matrix $[V_G\, \boldsymbol{R}](t)$ characterizes a sufficiently informative input-output signal pair at time t.

Remark 3.1.2 (Final) *Note that a scenario of sufficiently informative signals at time t for one of the described augmentation methods may eventually lead to undetermination in the other two approaches. Moreover, within the latter two frameworks, different choices of the kernels may lead to different sets of sufficiently informative signals. Therefore, the problems of identifying families of admissible kernels and of studying their properties (both in the deterministic and stochastic settings) will be dealt in the following.*

In order to emphasize the generality of the proposed methodology, we still have not assigned explicit analytic expressions to the kernels $K_1, \ldots, K_N$ and to $R_{N,j}$, $j \in p, \ldots, n+p$, which are needed to compute the auxiliary signals. The problem of identifying a class of non-asymptotic kernels yielding stable finite-dimensional state-space realizations will be addressed in the following sections.

3.2 Univariate Non-Asymptotic Kernels

In this section, we first consider a class of simplified univariate kernels $W(\tau), \tau \in \mathbb{R}_{\geq 0}$. The Volterra operators induced by this kind of kernels are given by

$$[V_W y](t) = \int_0^t W(\tau) y(\tau) d\tau,\ \ \forall t \in \mathbb{R}_{\geq 0}\,,$$

and are typically known as *weighted integral* operators. In our setting, we consider weighting patterns $W(\cdot) \in \mathcal{L}^2_{loc}(\mathbb{R}_{\geq 0})$ satisfying the non-asymptoticity conditions up to the i-th order and we call them *Univariate Non-asymptotic Kernels* (U-NKs).

Remark 3.2.1 (U-NKs Vs. Modulating Functions) *Within the class of IM identification methods, the MF approach uses time-dependent univariate kernels to minimize the effect of unknown initial conditions on the estimates. Univariate kernels, in our setting, need to fulfill weaker assumptions than those that the usual modulating functions need to subsume in the identification context (see [14]): while in the usual MF approach the width of the integration window is a critical parameter for noise sensitivity and the estimates are available only at the end of the integration interval, in our setting, a point-wise estimate (independent from initial conditions) is available at any time instant $t > 0$, and the integration process may proceed indefinitely without re-initialization. This is a distinctive feature of the proposed methodology.*

In order to exploit the U-NKs for parameter estimation, let us specialize the kernel construction procedure outlined in Theorem 2.1.3 to produce the needed causal auxiliary signals. To this end, consider $n+p$ modulating functions

$$W_1(\tau), W_2(\tau), \ldots, W_{n+p}(\tau)$$

and assume that W_1 is at least $(n+p-1)$-th order non-asymptotic (see Theorem 2.1.3). Now, analogously to (3.3), by applying the composed operator V_{P_N}, with $P_N = W_{n+p} \bullet \ldots \bullet W_p$, we obtain the transformed dynamic constraint:

$$[V_{P_N} y^{(n+p)}](t) = \sum_{i=0}^{n-1} a_i [V_{P_N} y^{(i+p)}](t) + \sum_{k=0}^{m-1} b_k [V_{P_N} u^{(k+p)}](t).$$

Thanks to Theorem 2.1.3, the surrogate signal derivatives in (3.5), in the U-NK setting, admit the following simple expressions:

$$\begin{aligned} r_{y,i}(t) &= [V_{P_N} y^{(i)}](\cdot), \quad i \in \{p, \ldots, n+p\}, \\ r_{u,i}(t) &= [V_{P_N} u^{(i)}](\cdot), \quad i \in \{p, \ldots, p+m-1\}. \end{aligned}$$

These signals can be used in place of the unmeasurable input-output derivatives to estimate the parameters.

Remark 3.2.2 *Admissible instances of U-NKs of the i-th order are, among many other possible functions:*

1. *the i-th order τ-monomial $W(\tau) = \tau^i$;*
2. *the damped unitary-step function $W(\tau) = (1 - e^{-\omega\tau})^i$;*
3. *the exp. damped i-th order τ-monomial $W(\tau) = \tau^i e^{-\omega\tau}$;*

where $\omega \in \mathbb{R}_{>0}$ is an arbitrary constant. According to (2.3), the first two kernels are locally square-integrable but not BIBO-stable, while the third one is BIBO-stable.

Now, it is worth noting that any BIBO stable univariate kernel $W(\cdot)$ has to satisfy the asymptotic condition $\lim_{\tau\to\infty} W(\tau) = 0$, in order to meet the requirement (2.3). Indeed, particularizing (2.3) to univariate kernels, we obtain

$$\sup_{t\in\mathbb{R}_{>0}} \int_0^t |W(\tau)|d\tau = \int_0^\infty |W(\tau)|d\tau < \infty .$$

In practice, it follows that the rate of update of the estimates generated by any BIBO-stable U-NK estimator decays as time proceeds far from the initial instant $t = 0$ because the weighting patterns fade toward zero. In other terms, the input-output injection undergoes an asymptotic suppression. This drawback will be addressed in the next section by using bivariate causal non-asymptotic kernels (still guaranteeing the internal stability of the estimator).

3.3 Bivariate Causal Non-Asymptotic Kernels

In this section, the main result is presented. To this end, let us introduce the following definition:

Definition 3.3.1 (i-th Order BC-NK) *If a kernel $K(\cdot,\cdot) \in \mathcal{HS}$, in addition to the assumptions posed in the statement of Lemma 2.1.1, for a given $i \geq 1$, verifies the conditions*

$$\left.\begin{array}{l} K^{(j)}(t,0) = 0 \\ K^{(j)}(t,t) = 0 \end{array}\right\} \begin{array}{l} \forall t \in \mathbb{R}_{\geq 0}, \\ \forall j \in \{0,\ldots,i-1\}, \end{array} \tag{3.13}$$

then, it is called an i-th Order Bivariate (strict) Causal Non-Asymptotic kernel. □

It is of customary importance to emphasize that only by using bivariate kernels all the conditions (3.13) can be fulfilled simultaneously. While Theorem 2.1.3 enabled us to construct auxiliary signals yielding the unavailable derivatives by taking advantage of non-asymptotic kernels, the following result can be used to exploit the causality property of BC-NKs to achieve the same task in an easier way. Indeed, by the conditions (3.13), thanks to Lemma 2.1.1, the image of a signal derivative can be expressed as

$$\left[V_{P_N} x^{(i)}\right](t) = (-1)^i \left[V_{P_N^{(i)}} x\right](t).$$

The following bivariate function is a possible instance of BC-NK, and it will be used in the sequel to carry out the design of the stable non-asymptotic estimator:

$$C_{\omega,N}(t,\tau) \triangleq e^{-\omega(t-\tau)}\left(1-e^{-\omega\tau}\right)^N\left(1-e^{-\omega(t-\tau)}\right)^N, \qquad (3.14)$$

where $\omega \in \mathbb{R}_{>0}$ is an arbitrary scalar parameter. The non-asymptoticity, causality and BIBO-stability properties of the devised kernel are illustrated by the following lemma.

Lemma 3.3.1 (Kernel Characterization: $C_{\omega,N}(t,\tau)$) *The bivariate kernel $C_{\omega,N}(t,\tau)$ is BIBO-stable and N-th order BC-NK. Moreover, all the kernel derivatives $C^{(i)}_{\omega,N}(t,\tau)$, with $i \in \{0,\ldots,N-1\}$, are BIBO-stable.* □

Proof 3 *First, we prove that the kernel $C_{\omega,N}(t,\tau)$ is a N-th order BC-NK. Indeed, all the non-anticipativity conditions up to the N-th order are met by the factor $(1-e^{-\omega\tau})^N$. The causality conditions up to the N-th order are met by the third factor $(1-e^{-\omega(t-\tau)})^N$. The BIBO-stability of $C^{(i)}_{\omega,N}(t,\tau)$ is implied by the fact that each $(e^{-\omega(t-\tau)})^{(i)}$, with $i \in \{0,\ldots,N-1\}$, is BIBO-stable and the following terms are bounded:$|(1-e^{-\omega(t-\tau)})^N| < 1$, $\forall\tau : 0 \le \tau \le t$ and $|(1-e^{-\omega(t-\tau)})^N| < 1$, $\forall\tau : 0 \le \tau \le t$ and their derivatives up to the $(N-1)$-th order are bounded too.*

Now, we describe how the image of the derivative $x^{(i)}(\cdot)$ through the operator $V_{C_{\omega,N}}$, *i.e.*, $\left[V_{C_{\omega,N}}x^{(i)}\right] = (-1)^i[V_{C^{(i)}_{\omega,N}}x]$ can be obtained as the output of a BIBO-stable finite-dimensional time-varying linear system.

First, the i-th derivative of the BC-NK (3.14) with respect to the second argument can be expressed as:

$$C^{(i)}_{\omega,N}(t,\tau) = \sum_{j=1}^{N+1} e^{-\omega j\,t} f_{\omega,N|i,j}(\tau) \qquad (3.15)$$

where $f_{\omega,N|i,j}(\cdot)$ are univariate functions of τ.
Let

$$C_{\omega,N|i,j}(t,\tau) \triangleq (-1)^i e^{-\omega j\,t} f_{\omega,N|i,j}(\tau)$$

Then, by the linearity of the Volterra operator, it follows that

$$[V_{C_{\omega,N}}x^{(i)}](t) = (-1)^i[V_{C^{(i)}_{\omega,N}}x](t) = \sum_{j=1}^{N+1}\left[V_{C_{\omega,N|i,j}}x\right](t). \qquad (3.16)$$

Moreover, letting $\xi_{i,j}(t) \triangleq \left[V_{C_{\omega,N|i,j}} x\right](t)$, with $i \in \{p, \ldots, n+p\}$, $j \in \{1, \ldots, N+1\}$, and considering that, $\forall t \in \mathbb{R}_{\geq 0}$:

$$\begin{array}{rcl} C_{\omega,N|i,j}(t,0) & = & 0 \\ \dfrac{\partial}{\partial t} C_{\omega,N|i,j}(t,\tau) & = & -\omega j \; e^{-\omega j\, t} \, f_{\omega,N|i,j}(\tau), \end{array} \tag{3.17}$$

then $\left[V_{C_{\omega,N}} x^{(i)}\right]$ admits the following $(N+1)$-th dimensional state-space realization:

$$\begin{cases} \xi_{i,j}^{(1)}(t) = C_{\omega,N|i,j}(t,t)x(t) - \omega j \, \xi_{i,j}(t), & \forall j \in \{1, \cdots, N+1\} \\ \left[V_{C_{\omega,N}} x^{(i)}\right](t) = \displaystyle\sum_{j=1}^{N+1} \xi_{i,j}(t), & \forall t \in \mathbb{R}_{\geq 0} \end{cases} \tag{3.18}$$
$$\xi_{i,1}(0) = 0, \ldots, \xi_{i,N+1}(0) = 0.$$

Being $|e^{-\omega j\, t} f_{\omega,N|i,j}(t)| < \infty$, $\forall j \in \{1, \ldots, N+1\}$, (*i.e.*, all the time-varying terms affine to the $x(t)$-injection are bounded), and since the system is diagonal with $\omega > 0$, then (3.18) is a BIBO-stable time-varying linear system.

Moreover, there exist finite scalars $\beta_{i,j} \in \mathbb{R}_{>0}$ such that

$$(e^{-\omega j\, t} f_{\omega,N|i,j}(t)) \overset{t \to \infty}{\longrightarrow} \beta_{i,j}$$

This implies that the time-varying system (3.18), for $t \to \infty$, tends a to stable linear time-invariant system in which the $x(t)$-injection is never suppressed. Thanks to (3.18), the extended auxiliary signal vector $\mathbf{z}_e(t)$, which embeds both the signals $z(t)$ and $r_{y,p+n}$ needed to form the constraint (3.6):

$$\mathbf{z}_e(t) = [r_{y,p}(t), \ldots, r_{y,p+n}(t), r_{u,p}(t), \ldots, r_{u,p+m-1}(t)],$$

with

$$\begin{array}{rcl} r_{y,i} & = & \left[V_{C_{\omega,N}} y^{(i)}\right], i \in \{p, \ldots, n+p\}, \\ r_{u,i} & = & \left[V_{C_{\omega,N}} u^{(i)}\right], i \in \{p, \ldots, m-1+p\}, \end{array}$$

can be obtained as the output of an overall $n_{\boldsymbol{\xi}} = (n+m+1)(N+1)$-dimensional linear time-varying dynamical system:

$$\mathcal{G}_{u,y \to \mathbf{z}_e} : \begin{cases} \boldsymbol{\xi}^{(1)}(t) = \mathbf{G}_\xi \boldsymbol{\xi}(t) + \mathbf{E}_y(t) y(t) + \mathbf{E}_u(t) u(t), \\ \mathbf{z}_e(t) \;\; = \mathbf{H}_\xi \boldsymbol{\xi}(t), & t \in \mathbb{R}_{\geq 0} \end{cases} \tag{3.19}$$
$$\boldsymbol{\xi}(0) = \mathbf{0},$$

where $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ is the overall state-vector,

$$
\mathbf{G} = \left[\begin{array}{ccc|ccc}
\mathbf{G}_p & & & & & \\
& \ddots & & & \mathbf{0} & \\
& & \mathbf{G}_{p+n} & & & \\
\hline
& & & \mathbf{G}_p & & \\
& \mathbf{0} & & & \ddots & \\
& & & & & \mathbf{G}_{p+m-1}
\end{array}\right] \in \mathbb{R}^{n_{\boldsymbol{\xi}} \times n_{\boldsymbol{\xi}}},
$$

$$
\mathbf{G}_i = \begin{bmatrix} -\omega & & 0 \\ & \ddots & \\ 0 & & -\omega(N+1) \end{bmatrix} \in \mathbb{R}^{(N+1)\times(N+1)},
$$

with $i \in \{p, \ldots, p+n\}$. Moreover, we have

$$
\mathbf{E}(t) = \left[\begin{array}{c} \mathbf{E}_p(t) \\ \vdots \\ \mathbf{E}_{p+n}(t) \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \in \mathbb{R}^{n_{\boldsymbol{\xi}}},\ \mathbf{F}(t) = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline \mathbf{F}_p(t) \\ \vdots \\ \mathbf{F}_{p+m-1}(t) \end{array}\right] \in \mathbb{R}^{n_{\boldsymbol{\xi}}},
$$

$$
\mathbf{E}_i(t) = \begin{bmatrix} C_{\omega,N|i,1}(t,t) \\ \vdots \\ C_{\omega,N|i,N+1}(t,t) \end{bmatrix} \in \mathbb{R}^{N+1}, \tag{3.20}
$$

$$
\mathbf{H} = \left[\begin{array}{ccc|ccc}
\mathbf{1}^T & & & & & \\
& \ddots & & & \mathbf{0} & \\
& & \mathbf{1}^T & & & \\
\hline
& & & \mathbf{1}^T & & \\
& \mathbf{0} & & & \ddots & \\
& & & & & \mathbf{1}^T
\end{array}\right] \in \mathbb{R}^{n_{\boldsymbol{\xi}} \times n_{\boldsymbol{\xi}}},
$$

where $\mathbf{1}^\top$ denotes a row vector of ones with $(N+1)$ elements.

By choosing the covariance filtering method as augmentation strategy (see Section 3.1), and by assuming that the V_G operator used for augmentation admits a one-dimensional stable state-space realization (take, for instance a kernel $G(t,\tau) = e^{-\omega(t-\tau)}$) the overall BC-NK estimator can be implemented as an *internally stable* $(n+m+1)(N+1) + (n+m+1)(n+m-1)$-th order linear time-varying dynamical system.

Indeed, the augmentation system can be viewed, in turn, as a CT dynamical system $\mathcal{A}_{\mathbf{z}_e \to \mathrm{vec}[\mathbf{Z},\mathbf{r}]}$, where $\mathrm{vec}[\mathbf{Z}, \mathbf{r}]$ represents the outcome of the augmentation, obtained by stacking the columns of $\mathbf{Z}(t)$ in a single vector to which $\mathbf{r}(t)$ is finally appended.

The dynamic part of the BC-NK estimator consists of the cascade of the auxiliary-signals-generation system $\mathcal{G}_{u,y \to \mathbf{z}_e}$ and of the augmentation system $\mathcal{A}_{\mathbf{z}_e \to \mathrm{vec}[\mathbf{Z},\mathbf{r}]}$. The internal stability of the BC-NK estimator refers to the stability of both these subsystems, but does not guarantee the boundedness of the estimates at time t. Indeed, according to Section 3.1, to obtain the parameter estimates we need a further processing step. The estimation process is completed by the algebraic inversion map (see Figure 3.1):

$$\mathcal{E}(\cdot) : \mathrm{vec}[\mathbf{Z}(t), \mathbf{r}(t)] \mapsto \hat{\boldsymbol{\theta}}(t)$$

where the estimated parameter vector $\hat{\boldsymbol{\theta}}(t)$ is obtained as

$$\hat{\boldsymbol{\theta}}(t) = \mathcal{E}\Big(\mathrm{vec}[\mathbf{Z}(t), \mathbf{r}(t)]\Big) \triangleq (\mathbf{Z}(t))^{\dagger}\, \mathbf{r}(t).$$

The estimation map $\mathcal{E}(\cdot)$ is not guaranteed to be bounded for all values of its argument, but only when the excitation condition outlined in Section 2.1.3 is met (the fulfillment of this condition depends on the informative content of the input-output signals restrictions $u_{(0,t)}(\cdot)$ and $y_{(0,t)}(\cdot)$). A supervision scheme can be introduced to check the invertibility of $\mathbf{Z}(t)$ in order to avoid singularities.

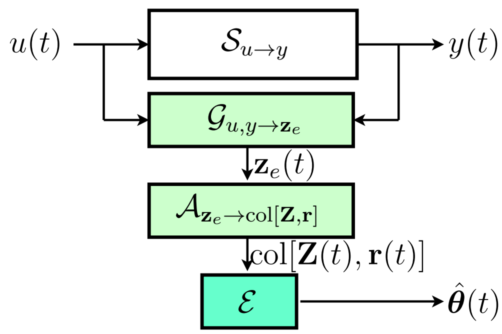


Fig. 3.1: Implementation block diagram. The dynamic part of the estimator consists in the cascade of the auxiliary signal generation system $\mathcal{G}_{u,y\to\mathbf{z}_e}$ followed by the augmentation system $\mathcal{A}_{\mathbf{z}_e\to\mathrm{vec}[\mathbf{Z},\mathbf{r}]}$. The estimated parameter vector $\hat{\boldsymbol{\theta}}$ is finally obtained by a static inversion map $\mathcal{E}$.

3.4 Concluding remarks

In this Chapter the Bivariate Causal Non-Asymptotic Kernels have been introduced, providing a solid theoretical results in order to obtain a fast and stable estimator. It has been derived the whole parametric estimation architecture and the convergence of the estimator has been proved. In the following Chapter, some simulation results will be presented in order to show the effectiveness of the proposed architecture.

Chapter 4

Simulation results

The identification methods for linear systems that were described in previous chapter are now applied to several examples to show the effectiveness of the proposed estimator and to compare the perforamces with estimators available in literature.

4.1 Mass-Spring-Damper example

Let us considerthe 1-DOF mass-spring-damper system model depicted in Figure 4.1; such a system consists of an inertial mass $M = 1\,kg$, a spring with elastic constant $k = 3\,Nm^{-1}$ and a linear damping element with $c = 2\,Ns\,m^{-1}$.

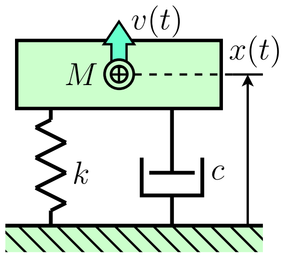


Fig. 4.1: Scheme of the 1-DOF mass-spring-damper system simulated in the example.

Using Newton's law of motion, we can derive the second-order differential equation for displacement x as a function of time:

$$\begin{cases} Mx^{(2)}(t) + cx^{(1)}(t) + kx(t) = v(t), & \\ y(t) = x(t) + o + \eta_y(t), & t \in \mathbb{R}_{\geq 0}, \\ u(t) = v(t), & \end{cases}$$
$$x(0) = x_0, \quad x^{(1)}(0) = x_0^{(1)}. \tag{4.1}$$

where $v(\cdot)$ represents a measurable external force-input for the system, $y(\cdot)$ is the measured position signal, affected by a constant measurement bias

$o = 1\,m$ and by an unstructured perturbation term $\eta_y(\cdot)$ (addressed to as *output measurement noise*), while $u(\cdot)$ is the measured forcing input signal. Neglecting the effect of unstructured perturbations $\eta_y(\cdot)$, the following input-output dynamic constraint can be obtained by rearranging (4.1):

$$y^{(2)}(t) = a_1 y^{(1)}(t) + a_0(y(t) - o) + b_0 u(t), \tag{4.2}$$

where $a_0 = -k\,M^{-1} = -3$, $a_1 = -c\,M^{-1} = -2$, $b_0 = M^{-1} = 1$. Now, to estimate the parameters in the presence of bias on the measurements, let us set $p = 1$ and, being $n = 2$ for the considered system, let us choose $N = 3$. The following U-NKs will yield the estimator described in [89]:

$$W_1(\tau) = \tau^3; \quad W_2(\tau) = H(\tau);$$
$$W_3(\tau) = H(\tau); \quad W_4(\tau) = H(\tau).$$

In view of Theorem 2.1.3 we can compute the auxiliary signals $r_{y,i}(\cdot)$ and $r_{u,i}(\cdot)$ as

$$\begin{aligned}
r_{x_1}(t) &= R_{3,1}(t,t)x(t) - R_{3,1}(t,0)x(0) - [V_{R^{(1)}_{3,1}} x](t),\\
r_{x_2}(t) &= R_{3,2}(t,t)x(t) - R_{3,2}(t,0)x(0) - [V_{R^{(1)}_{3,2}} x](t),\\
r_{x_3}(t) &= R_{3,3}(t,t)x(t) - R_{3,3}(t,0)x(0) - [V_{R^{(1)}_{3,3}} x](t),
\end{aligned}$$

with $x \leftarrow \{u, y\}$ and where the R-kernels are obtained from the U-NKs W_1, W_2 and W_3 by the iterative procedure outlined in the statement of Theorem 2.1.3. For the chosen U-NKs it holds that

$$\begin{aligned}
R_{3,1}(t,t) &= 0, & R_{3,1}(t,0) &= 0,\\
R_{3,2}(t,t) &= 0, & R_{3,2}(t,0) &= 0,\\
R_{3,3}(t,t) &= -t^3, & R_{3,3}(t,0) &= 0,
\end{aligned}$$

and

$$\begin{aligned}
R^{(1)}_{3,1}(t,\tau) &= -\frac{3t^2\tau^2}{2} + 4t\tau^3 - \frac{5\tau^4}{2},\\
R^{(1)}_{3,2}(t,\tau) &= 3t^2\tau - 12t\tau^2 + 10\tau^3,\\
R^{(1)}_{3,3}(t,\tau) &= -3t^2 + 24t\tau - 30\tau^2,
\end{aligned}$$

yielding the integral forms

$$\begin{aligned}
r_{x_1}(t) &= \frac{3}{2}t^2\int_0^t \tau^2 x(\tau)d\tau - 4t\int_0^t \tau^3 x(\tau)d\tau + \frac{5}{2}\int_0^t \tau^4 x(\tau)d\tau,\\
r_{x_2}(t) &= -3t^2\int_0^t \tau x(\tau)d\tau + 12t\int_0^t \tau^2 x(\tau)d\tau - 10\int_0^t \tau^3 x(\tau)d\tau,\\
r_{x_3}(t) &= -t^3 x(t) - 3t^2\int_0^t x(t)d\tau - 24t\int_0^t \tau x(\tau)d\tau + 30\int_0^t \tau^2 x(\tau)d\tau.
\end{aligned} \tag{4.3}$$

Note that, while in [89] the auxiliary signals are expressed in terms of nested integrals, here we have reported the equivalent single-integral expressions, returning the same signals. The interested reader can obtain the expressions in (4.3) directly from from the nested integrals forms presented in [89] by considering the following Volterra kernel composition result (see equations (8.4) and (8.5) in the Appendix):

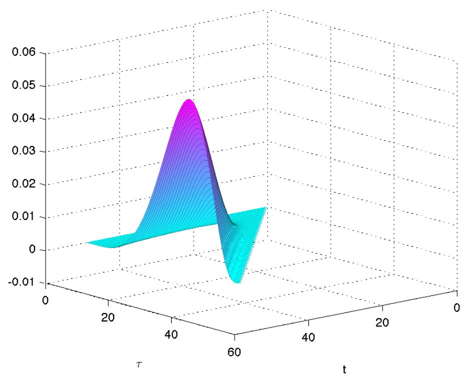
$$\underbrace{H(\tau) \bullet \ldots \bullet H(\tau)}_{i-1 \text{ times}} \bullet \tau^j = \frac{\tau^j (t-\tau)^{i-1}}{i!}, \ \forall j \in \mathbb{Z},$$

which implies that

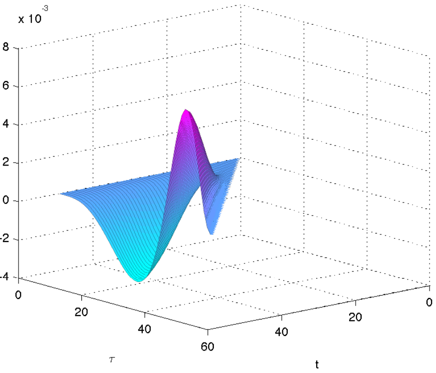
$$\begin{aligned}
\int_0^t \int_0^{\sigma_1} \tau^j x(\tau) d\tau d\sigma_1 \quad &= \left[V_{H \bullet \tau^j} \ x\right](t), \\
&= \int_0^t \tau^j (t-\tau) x(\tau) d\tau, \\
\int_0^t \int_0^{\sigma_2} \int_0^{\sigma_1} \tau^j x(\tau) d\tau d\sigma_1 d\sigma_2 \quad &= \left[V_{H \bullet H \bullet \tau^j} \ x\right](t), \\
&= \int_0^t \frac{\tau^j (t-\tau)^2}{2} x(\tau) d\tau.
\end{aligned}$$

Successive integration has ben used in this case as augmentation method. By chosing $G = H(\tau)$ as kernel of the augmentation operator V_G, then U-NK estimator will exactly reproduce the estimator in [89].

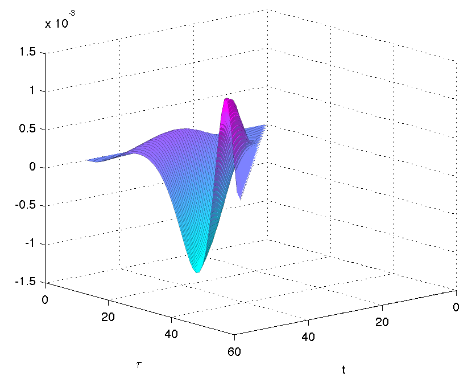
To carry out a comparable simulation in the BC-NK framework, the same $N = 3$ value has been used in the implementation of the BC-NK kernel (3.14). The kernel parameter ω has been set to $\omega = 1$ (this is an arbitrary choice; some considerations in order to design the value of the kernel parameter will be presented in Chapter 6). The procedure for constructing the auxiliary signals generation system by BC-NK kernels (whose representations are depicted in Figure 4.2) consists in taking the derivatives $C^{(i)}_{\omega,N}(t,\tau), \ \ i \in \{1,2,3\}$ of the BC-NK (3.14), then in identifying the terms $C_{\omega,N|i,j}$, with $j \in \{1,2,3,4\}$ (see (3.15) and (3.16)), and finally in computing $C_{\omega,N|i,j}(t,t)$ to form the $\mathbf{E}_i(t)$ matrices (see (3.20)) needed for the implementation of the auxiliary signal generation system $\mathcal{G}_{u,y\to\mathbf{z}_e}$ (see (3.19)). Neglecting the intermediate algebraic manipulations, we get:



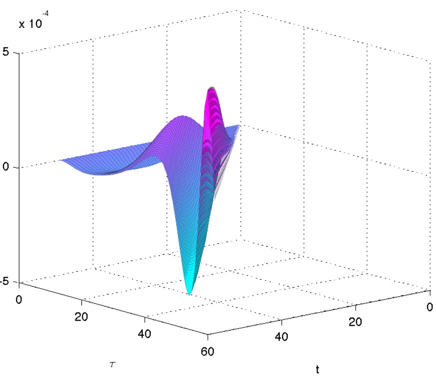
(a) $C_{\omega,N}(t,\tau)$



(b) $C^{(1)}_{\omega,N}(t,\tau)$



(c) $C^{(2)}_{\omega,N}(t,\tau)$



(d) $C^{(3)}_{\omega,N}(t,\tau)$

Fig. 4.2: Graphical rappresentation of the Bivariate Causal Non-Asymptotic Kernel (3.14) and its derivatives (see (3.15)), with $\omega = 0.1$.

$$\mathbf{E}_1 = \begin{bmatrix} -(2e^{-t}+1)(e^{-t}-1)^2 \\ 3(e^{-t}-1)^2(e^{-t}+2) \\ -9(e^{-t}-1)^2 \\ -(e^{-t}-1)^2(e^{-t}-4) \end{bmatrix},$$

$$\mathbf{E}_2 = \begin{bmatrix} -4e^{-3t}+3e^{-2t}+1 \\ 3e^{-3t}+9e^{-t}-12 \\ 9(e^{-t}-1)(e^{-t}-3) \\ (e^{-t}-1)(e^{-2t}-11e^{-t}+16) \end{bmatrix},$$

$$\mathbf{E}_3 = \begin{bmatrix} -8e^{-3t}+3e^{-2t}-1 \\ 3e^{-3t}-9e^{-t}+24 \\ -9e^{-2t}+72e^{-t}-81 \\ -e^{-3t}+24e^{-2t}-81e^{-t}+64 \end{bmatrix},$$

while

$$\mathbf{G}_i = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -4 \end{bmatrix}, \quad i \in \{1,2,3\}.$$

Covariance filtering has been used in this case as augmentation method. The kernel G of the filtering operator V_G has been chosen as $G(t,\tau) = e^{-0.1(t-\tau)}$, which yields as a single-dimensional stable linear system. The construction of the augmentation system with the chosen kernel is trivial, so its state-space realization is omitted for brevity.

4.1.1 Non-Asymptotic Parameter Estimation in the Noise-free Scenario

The first simulation deals with a noise-free scenario, *i.e.*, $\eta_u(t) = 0, \, \; \forall t \in \mathbb{R}_{\geq 0}$.

The initial conditions for the mass-spring-damper system have been set to $x(0) = 1\,m$ and $x^{(1)}(0) = 10\,ms^{-1}$, while the forcing input has been chosen as a sum of sinusoids:

$$v(t) = 10\sin(t) + \sin(10\,t).$$

Figure 4.3 depicts the measured input and output signals $u(t) = v(t)$ and $y(t) = x(t) + o$ in the noise-free case. Although the theoretical instantaneity of the method gets lost in the time-discretization of the estimator's dynamics, in the digital representation of the signals and in the numerical computation of the pseudo-inverse, the parameters are correctly esimated with negligible duration of the transient by both methods, as shown in Figure 4.4. Remarkably, the proposed BC-NK estimator combines the practical instantaneity with the internal stability, so that re-initialization is not required. In Figures 4.5 and 4.6, the time-behaviors of the singular values $\Sigma(\mathbf{Z}(t))$ of the $\mathbf{Z}(t)$ matrices yielded by the augmentation systems of the two estimators are shown. As can be observed, the BC-NK technique shows a bounded behavior of the singular values $\Sigma(\mathbf{Z}(t))$, whereas the other estimation technique, though showing a fast convergence behavior toward accurate estimates of the parameters, requires periodic reset in order to cope with the integrator windup issue.

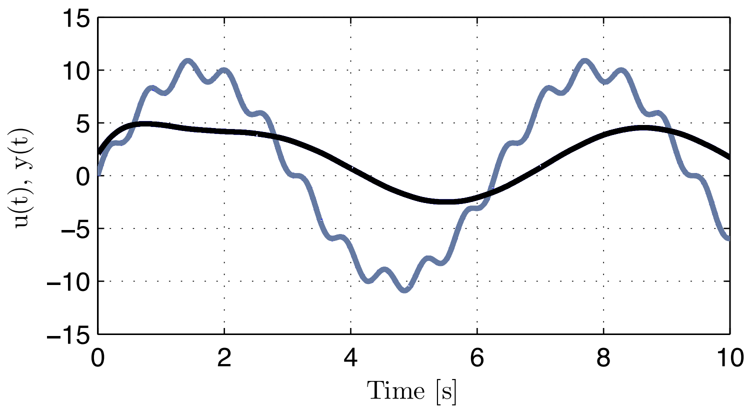


Fig. 4.3: Trends of the measured signals $u(t)$ (gray) and $y(t)$ (black) used for the estimation in noise-free conditions.

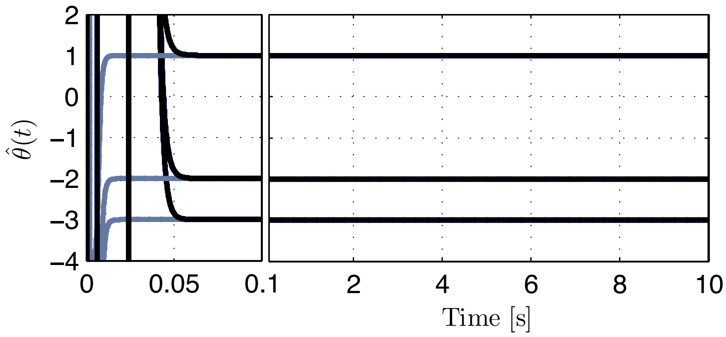


Fig. 4.4: The parameters estimated by the U-NK method (gray) and by the BC-NK estimator (black) in noise-free conditions converge to the true values $a_0 = -3$, $a_1 = -2$ and $b_0 = 1$. The initial part of the simulation has been magnified to show the fast convergence of both methods. The theoretical instantaneity cannot be achieved because of time-discretization and numerical precision issues.

It is worth noting that no high-gain output injection has been performed by the two methods. In this respect, a further simulation has been carried

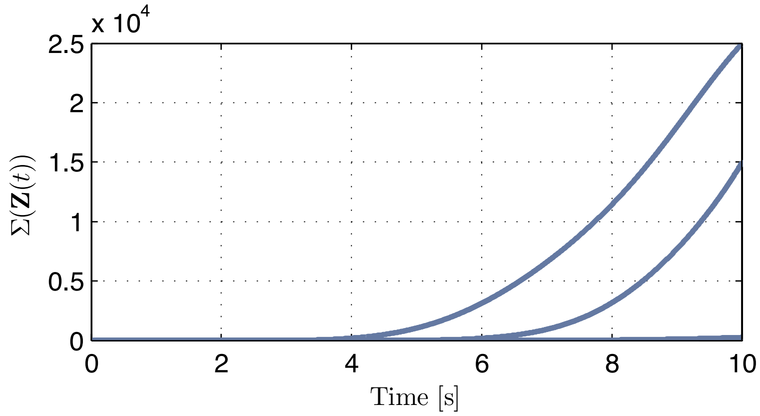


Fig. 4.5: Singular values of the $\mathbf{Z}(t)$ matrix produced by the U-NK estimator with successive integration. The unstable integrators used to generate the auxiliary signals lead the singular values of $\mathbf{Z}(t)$ to diverge. Periodic reset of the estimator is needed to avoid numerical windup.

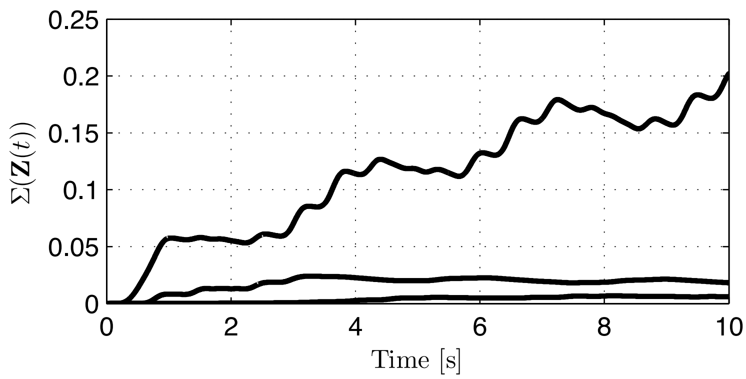


Fig. 4.6: Singular values $\Sigma(\mathbf{Z}(t))$of the $\mathbf{Z}(t)$ matrix produced by the BC-NK estimator with covariance filtering. The internal stability of the estimator guarantees that the singular values of $\mathbf{Z}(t)$ will remain bounded for any bounded input-output signal pair.

out in noisy conditions.

4.1.2 Estimation with Unstructured Measurement Perturbations

Let's consider a different scenario, where the additive output measurement noise $\eta_y(\cdot)$ has been simulated as uniformly distributed random signals taking values respectively in the intervals $[-0.8, 0.8]$. The perturbed signal used for parameter estimation are depicted in Figure 4.7.

As can be seen from Figure 4.8 the BC-NK estimator shows good noise immunity and the estimated parameters converge to a neighborhood of the true values. Conversely, the U-NK estimator, implemented without further provision for removing the noise effect, does not provide satisfactory results; in this respect, as suggested in [89], it possible to further process the estimates by a low-pass filter in order to mitigate the influence of noise.

Remarkably, although the BC-NK method has shown to be inherently robust in front of measurement perturbations, further improvements can be obtained by tuning the ω parameter of the non-asymptotic causal kernel (3.14) and by accurately choosing the filtering operator V_G used for augmentation. The point now is characterize, in both deterministic and stochastic settings, the behavior of the BC-NK estimator in the presence of measurement noise, in order to determine tuning rules for the aforementioned parameters (this argument will be dealt in Chapter 6).

4.2 Benchmark Example

In this section, we compare by numerical simulations the BC-NK estimator with the U-NK estimator, considering a benchmark proposed by Rao and Garnier in [90] (see also [91] and [92]): it is a 4-th order, non minimum-phase system whose dynamics are described by:

$$\begin{cases} x^{(4)}(t) = a_1 x^{(3)}(t) + a_2 x^{(2)}(t) + a_3 x^{(1)}(t) + a_4 x(t) + b_1 u^{(1)}(t) + b_2 u(t), \\ y(t) = x(t) + o + \eta_y(t), \\ u(t) = v(t), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t \in \mathbb{R}_{\geq 0}, \end{cases}$$
$$x(0) = x_0,\ \ x^{(1)}(0) = x_0^{(1)},\ \ x^{(2)}(0) = x_0^{(2)},\ \ x^{(3)}(0) = x_0^{(3)}, \tag{4.4}$$

where $b_1 = -6400$, $b_2 = 1600$, $a_1 = 5$, $a_2 = 408$, $a_3 = 416$ and $a_4 = 1600$, while $y(\cdot)$ is the output of the system affected by the *output measurement*

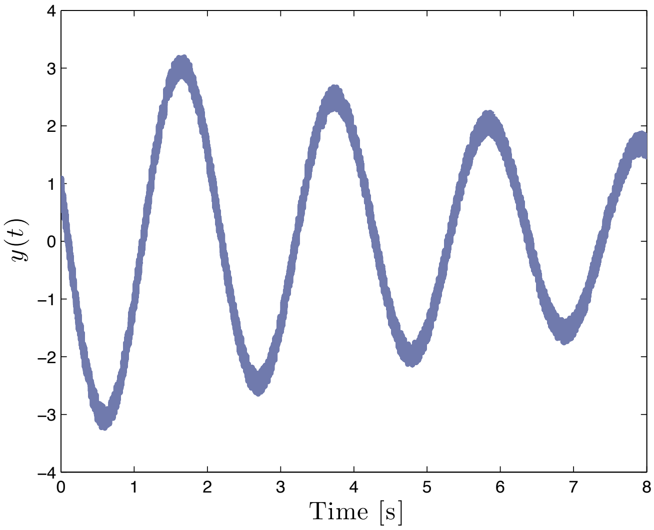


Fig. 4.7: Trends of the noisy output measurement signal $y(t)$ used for the estimation.

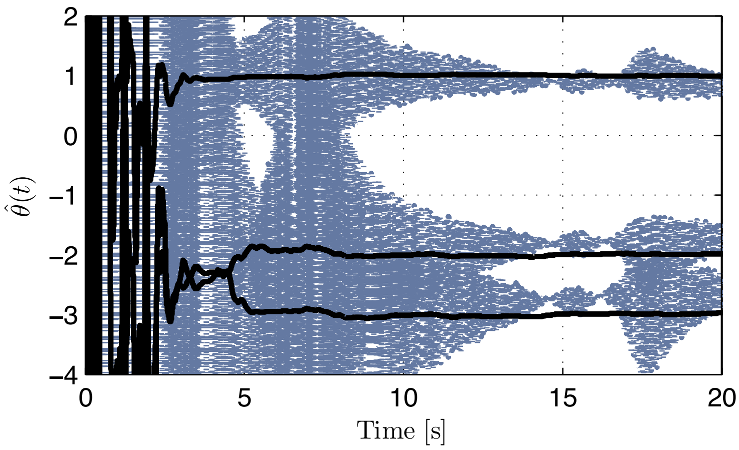


Fig. 4.8: Parameters estimated by the U-NK method (gray) and by the BC-NK estimator (black) with noisy measurements. The BC-NK estimator with covariance filtering provides reliable estimates even in presence of noise.

noise $\eta_y(\cdot)$ and a bias $o = 1$, while $u(\cdot)$ is the measured forcing input signal. Neglecting the effect of unstructured perturbations $\eta_u(\cdot)$, the following transfer function describes the input-output behavior of (4.4):

$$G(s) = \frac{k_1 s + k_2}{s^4 + c_1 s^3 + c_2 s^2 + c_3 s + c_4} \tag{4.5}$$

where $c_1 = -a_1 = -5$, $c_2 = -a_2 = -408$, $c_3 = -a_3 = -416$,
$c_4 = -a_4 = -1600$, $k_1 = b_1 = -6400$, $k_2 = b_2 = 1600$. Now, to estimate the parameters in the presence of bias on the measurements, let us set $p = 1$ and, being $n = 4$ for the considered system, let us choose $N = 5$ and this will yeld to the following U-NKs:

$$\begin{aligned} W_1(\tau) &= \tau^3; \quad W_2(\tau) = H(\tau); \\ W_3(\tau) &= H(\tau); \quad W_4(\tau) = H(\tau); \\ W_5(\tau) &= H(\tau). \end{aligned}$$

Thanks of Theorem 2.1.3 we can compute the auxiliary signals $r_{y,i}(\cdot)$ and $r_{u,i}(\cdot)$ as

$$r_{x_i}(t) \quad = R_{5,i}(t,t)x(t) - R_{5,i}(t,0)x(0) - [V_{R^{(1)}_{5,i}} x](t), \qquad \forall i \in \{1, \dots, 5\}$$

with $x \leftarrow \{u, y\}$ and where the R-kernels are obtained from the U-NKs W_1, W_2 and W_3. For the chosen U-NKs it holds that

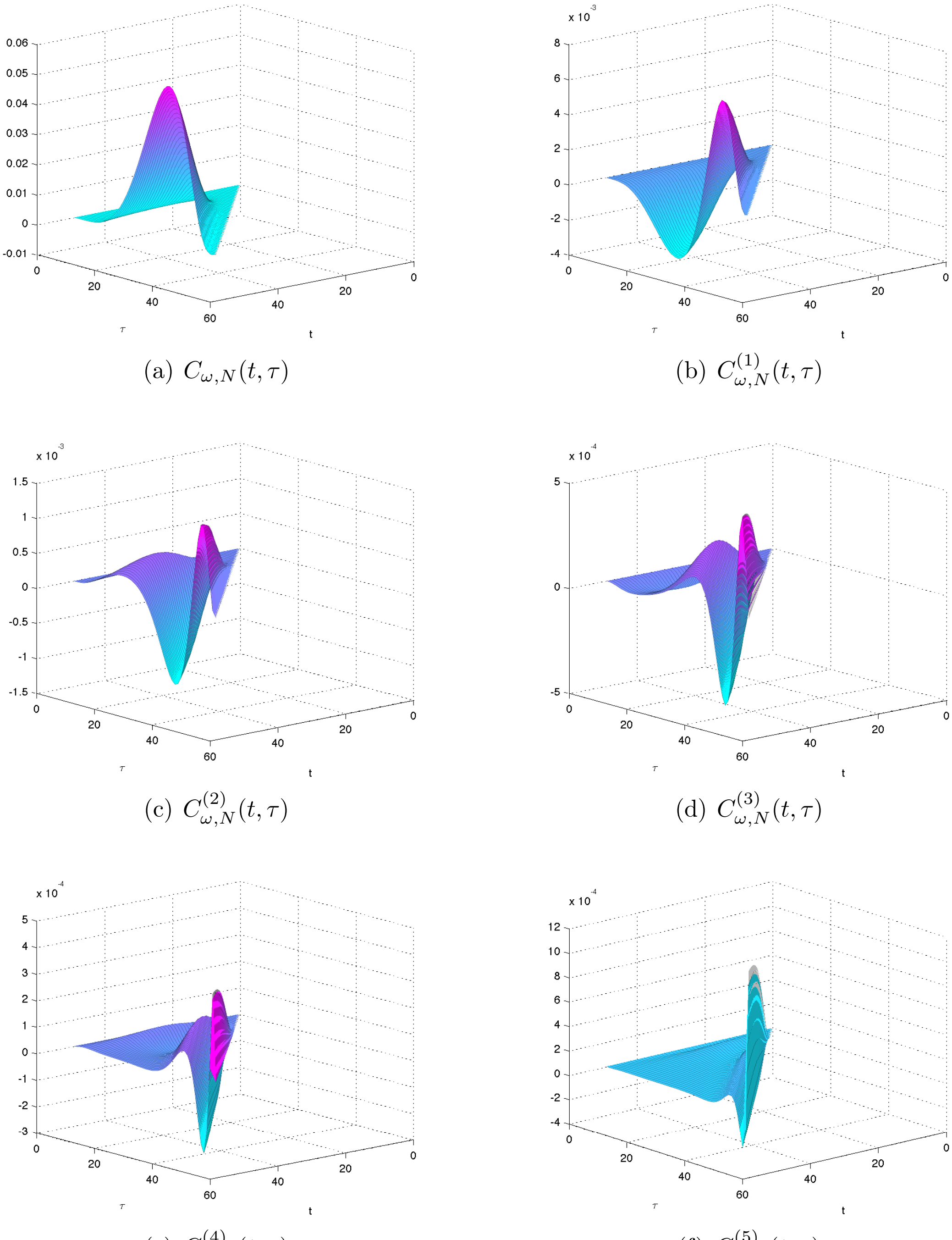
$$R_{5,i}(t,t) = 0 \;\; R_{5,i}(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}, \forall i \in \{1, \dots, 5\}$$

and

$$\begin{aligned} R^{(1)}_{5,1}(t,\tau) &= -(5t^4\tau^4)/24 + t^3\tau^5 - (7t^2\tau^6)/4 + (4t\tau^7)/3 - (3\tau^8)/8, \\ R^{(1)}_{5,2}(t,\tau) &= 5t^4\tau^3/6 - 5t^3\tau^4 + 21t^2\tau^5/2 - (28t\tau^6)/3 + 3\tau^7, \\ R^{(1)}_{5,3}(t,\tau) &= -(5t^4\tau^2)/2 + 20t^3\tau^3 - (105t^2\tau^4)/2 + 56t\tau^5 - 21\tau^6, \\ R^{(1)}_{5,4}(t,\tau) &= 5t^4\tau - 60t^3\tau^2 + 210t^2\tau^3 - 280t\tau^4 + 126\tau^5, \\ R^{(1)}_{5,5}(t,\tau) &= -5t^4 + 120t^3\tau - 630t^2\tau^2 + 1120t\tau^3 - 630\tau^4. \end{aligned}$$

yielding the integral forms

$$\begin{aligned} r_{x_1}(t) &= \int_0^t \Big(-\frac{5}{24}t^4\tau^4 + t^3\tau^5 - \frac{7}{2}t^2\tau^6 + \frac{4}{3}t\tau^7 - \frac{3}{8}\tau^8 \Big) x(\tau)d\tau, \\ r_{x_2}(t) &= \int_0^t \Big(\frac{5}{6}t^4\tau^3 - 5t^3\tau^4 + \frac{21}{2}t^2\tau^5 - \frac{28}{3}t\tau^6 + 3\tau^7 \Big) x(\tau)d\tau, \\ r_{x_3}(t) &= \int_0^t \Big(-\frac{5}{2}t^4\tau^2 + 20t^3\tau^3 - \frac{105}{2}t^2\tau^4 56t\tau^5 - 21\tau^6 \Big) x(\tau)d\tau, \\ r_{x_4}(t) &= \int_0^t \Big(5t^4\tau - 60t^3\tau^2 + 210t^2\tau^3 - 280t\tau^4 + 126\tau^5 \Big) x(\tau)d\tau, \\ r_{x_5}(t) &= -5t^4 x(t) + \int_0^t \Big(120t^3\tau - 630t^2\tau^2 + 1120t\tau^3 - 630\tau^4 \Big) x(\tau)d\tau. \end{aligned}$$



(a) $C_{\omega,N}(t,\tau)$

(b) $C^{(1)}_{\omega,N}(t,\tau)$

(c) $C^{(2)}_{\omega,N}(t,\tau)$

(d) $C^{(3)}_{\omega,N}(t,\tau)$

(e) $C^{(4)}_{\omega,N}(t,\tau)$

(f) $C^{(5)}_{\omega,N}(t,\tau)$

Fig. 4.9: Graphical rappresentation of the Bivariate Causal Non-Asymptotic Kernel (3.14) and its derivatives (see (3.15)), with $\omega = 0.1$.

As in the previous section, to carry out a comparable simulation in the BC-NK framework, the same $N = 5$ value has been used in the implementation of the BC-NK kernels (whose representations are depicted in Figure 4.9). The kernel parameter ω has been set to $\omega = 1$. The procedure for constructing the auxiliary signals generation is the same shown in Section 4.1. Neglecting the intermediate algebraic manipulations, we get:

$$\mathbf{E}_1 = \begin{bmatrix} -(4e^{-t}+1)(e^{-t}-1)^4 \\ 5(3e^{-t}+2)(e^{-t}-1)^4 \\ -10(2e^{-t}+3)(e^{-t}-1)^4 \\ 10(e^{-t}-1)^4(e^{-t}+4) \\ -25(e^{-t}-1)^4 \\ -(e^{-t}-1)^4(e^{-t}-6) \end{bmatrix},$$

$$\mathbf{E}_2 = \begin{bmatrix} -(e^{-t}-1)^3(16e^{-t2}+3e^{-t}+1) \\ 5(e^{-t}-1)^3(9e^{-t2}+7e^{-t}+4) \\ -10(e^{-t}-1)^3(4e^{-t2}+7e^{-t}+9) \\ 10(e^{-t}-1)^3(e^{-t2}+3e^{-t}+16) \\ 25(e^{-t}-1)^3(e^{-t}-5) \\ (e^{-t}-1)^3(e^{-t2}-17e^{-t}+36) \end{bmatrix},$$

$$\mathbf{E}_3 = \begin{bmatrix} -(e^{-t}-1)^2(64e^{-t3}-7e^{-t2}+2e^{-t}+1) \\ 5(e^{-t}-1)^2(27e^{-t3}+14e^{-t2}+11e^{-t}+8) \\ -10(e^{-t}-1)^2(8e^{-t3}+11e^{-t2}+14e^{-t}+27) \\ 10(e^{-t}-1)^2(e^{-t3}+2e^{-t2}-7e^{-t}+64) \\ -25(e^{-t}-1)^2(e^{-t2}-14e^{-t}+25) \\ -(e^{-t}-1)^2(e^{-t3}-38e^{-t2}+193e^{-t}-216) \end{bmatrix},$$

$$\mathbf{E}_4 = \begin{bmatrix} -256e^{-t5}+405e^{-t4}-160e^{-t3}+10e^{-t2}+1 \\ 405e^{-t5}-400e^{-t4}+50e^{-t3}+25e^{-t}-80 \\ -160e^{-t5}+50e^{-t4}+100e^{-t2}-800e^{-t}+810 \\ 10e^{-t5}+100e^{-t3}-1600e^{-t2}+4050e^{-t}-2560 \\ 25(e^{-t}-1)(e^{-t3}-31e^{-t2}+131e^{-t}-125) \\ (e^{-t}-1)(e^{-t4}-79e^{-t3}+731e^{-t2}-1829e^{-t}+1296) \end{bmatrix},$$

$$\mathbf{E}_5 = \begin{bmatrix} -1024e^{-t5}+1215e^{-t4}-320e^{-t3}+10e^{-t2}-1 \\ 1215e^{-t5}-800e^{-t4}+50e^{-t3}-25e^{-t}+160 \\ -320e^{-t5}+50e^{-t4}-100e^{-t2}+1600e^{-t}-2430 \\ 10e^{-t5}-100e^{-t3}+3200e^{-t2}-12150e^{-t}+10240 \\ -25e^{-t4}+1600e^{-t3}-12150e^{-t2}+25600e^{-t}-15625 \\ -e^{-t5}+160e^{-t4}-2430e^{-t3}+10240e^{-t2}-15625e^{-t}+7776 \end{bmatrix},$$

while:

$$\mathbf{G} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 & -6 \end{bmatrix}.$$

Covariance filtering has been used in this case as augmentation method. The kernel G of the filtering operator V_G has been chosen as $G(t, \tau) = e^{-0.1(t-\tau)}$.

4.2.1 Non-Asymptotic Parameter Estimation in the Noise-free Scenario

The first simulation deals with a noise-free scenario, *i.e.*, $\eta_y(t) = 0$, $\forall t \in \mathbb{R}_{\geq 0}$. The initial conditions for system (4.5) have been set to $x(0) = 1$, $x^{(1)}(0) = 10$ and $x^{(2)}(0) = x^{(3)}(0) = 0$, while the forcing input has been chosen as a sum of sinusoids

$$v(t) = 10\sin(5t) + 6\sin(20t) + 3\sin(8t) + \sin(2t) + 7\sin(4t) + 9\sin(12t),$$

depicted in Figure 4.10.

Although the theoretical instantaneity of the method gets lost in the time-discretization of the estimator's dynamics, in the digital representation of the signals and in the numerical computation of the pseudo-inverse, the parameters have been correctly esimated with negligible delay by both methods, as shown in Figure 4.11. It is worth noting that the proposed BC-NK estimator, beyond fast convergence,is characterized by guaranteed internal stability, hence re-initialization is not required. In Figures 4.12 and 4.13, we show the time-behaviors of the singular values $\Sigma(\mathbf{Z}(t))$ of the $\mathbf{Z}(t)$ matrix produced by the two augmentation systems. As can be observed, the BC-NK technique yelds a bounded behavior of the singular values $\Sigma(\mathbf{Z}(t))$, wheras the U-NK method requires periodic reset in order to cope with the integrator windup issue.

4.2.2 Estimation with Unstructured Measurement Perturbations

In this example, the additive output measurement noise $\eta_y(\cdot)$ has been simulated as a uniformly distributed random signal taking values in the

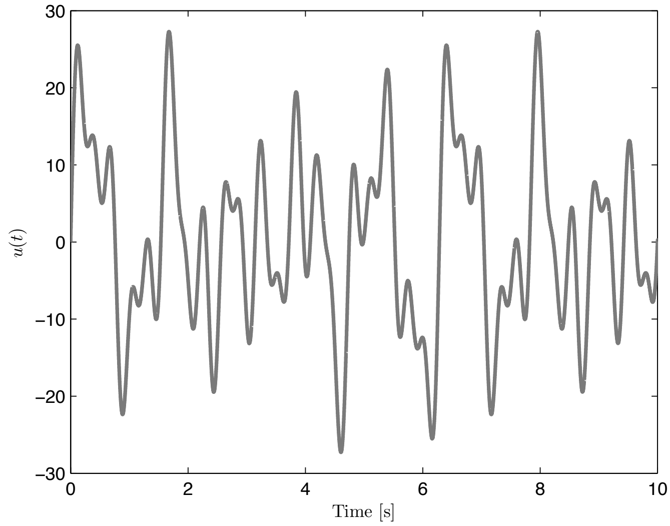
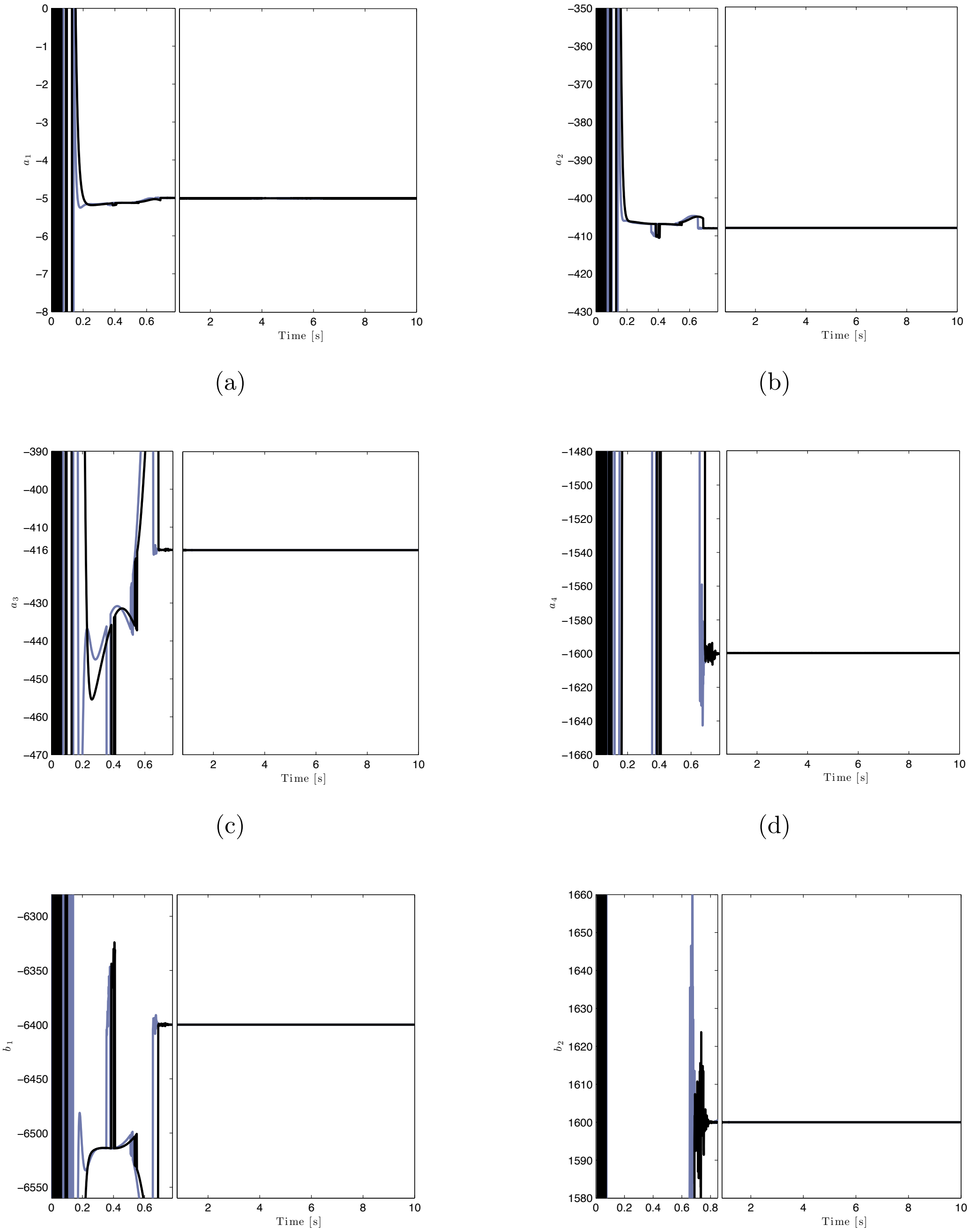


Fig. 4.10: Trends of the measured signals $u(t)$ used for the estimation in noise-free conditions.

interval $[-0.5, 0.5]$. The perturbed signal used for parameter estimation is depicted in Fig. 4.14.

As can be seen in Fig. 4.15, the BC-NK estimator shows good robustness against the output noise. Clearly, in noisy-conditions, it is important to have a knowledge of the bias in the computed noisy estimate, to this end we led the bias computation presented in Chapter 6.

In the following there will be presented some results in order to compare the parametric estimator proposed with two estimation techniques available in literature.



(a) (b)

(c) (d)

(e) (f)

Fig. 4.11: The parameters estimated by the U-NK method (gray) and by the BC-NK estimator (black) in noise-free conditions converge to the true values $b_1 = -6400$, $b_2 = 1600$, $a_1 = -5$, $a_2 = -408$, $a_3 = -416$ and $a_4 = -1600$.

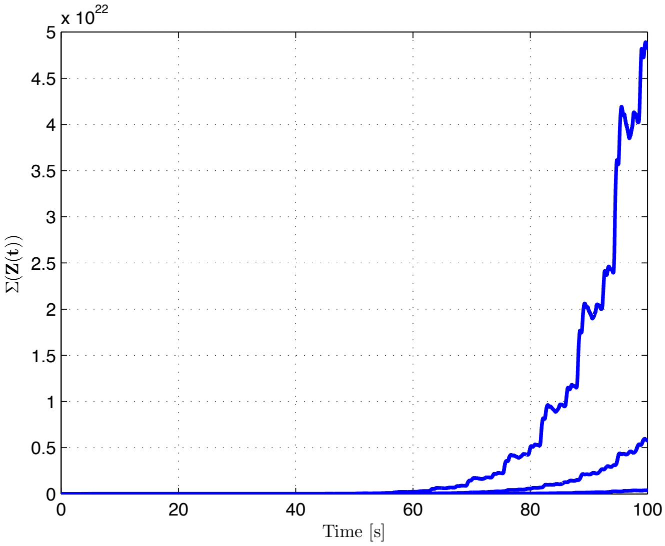


Fig. 4.12: Singular values of the $\mathbf{Z}(t)$ matrix produced by the U-NK estimator. The non-BIBO kernels used to generate the auxiliary signals lead the singular values of $\mathbf{Z}(t)$ to diverge. Periodic reset of the estimator is needed to avoid numerical windup.

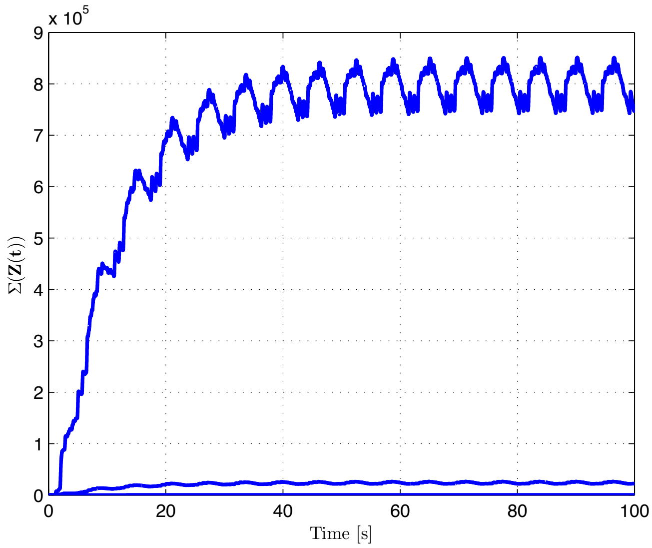


Fig. 4.13: Singular values $\Sigma(\mathbf{Z}(t))$of the $\mathbf{Z}(t)$ matrix produced by the BC-NK estimator with covariance filtering. The internal stability of the estimator guarantees that the singular values of $\mathbf{Z}(t)$ will remain bounded for any bounded input-output signal pair.

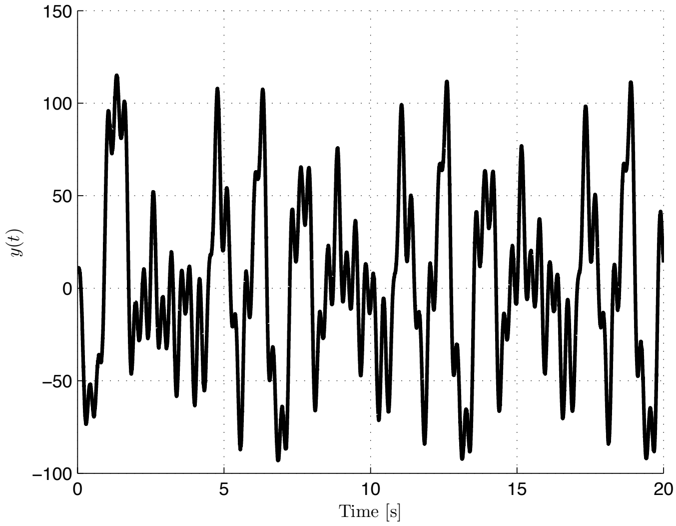
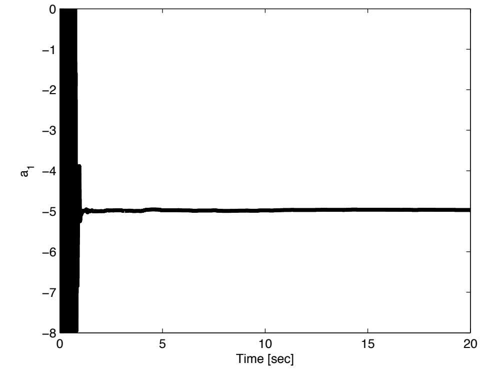


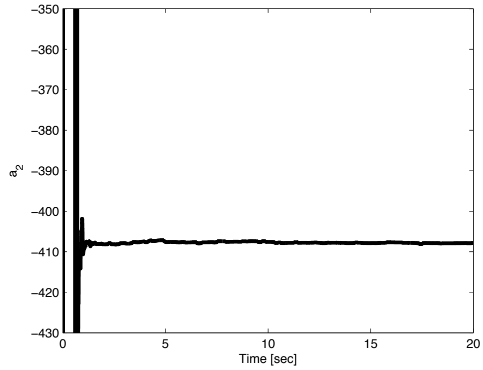
Fig. 4.14: Trends of the measured signals $y(t)$ used for the estimation in noisy conditions.

4.3 Comparison of the BC-NK method

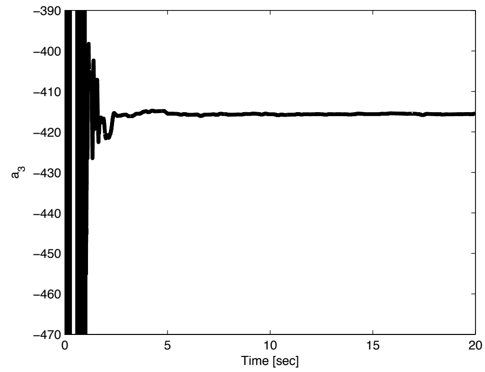
What follows are different simulations experiments, containing Monte-Carlo simulations to compare BC-NK parameter estimation method discussed in Chapter 3 with two methods falling into the two approaches discussed in Section 1.2.1, *i.e.*, State Variable Filtering (SVF) and Integral Methods (IM). In both of these cases the aim is to low pass filter the data in order to: *i*) get a stable estimate of the derivatives, and *ii*) avoid a strong gain in the high frequencies where the signal is mostly composed of noise.
Therefore, it has been comparised the BC-NK method with an integral method present in literature,*i.e.*, Hartley Modulating Function (HMF, see *e.g.*, [93] and [28])) and with a SVF method, *i.e.*, the Refined Instrumental Variable Method (SRIVC, see *e.g.*, [65] and [94]).



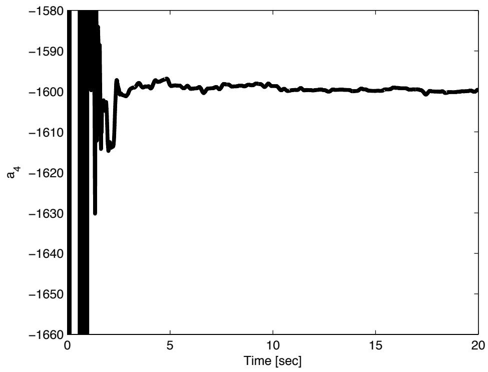
(a)



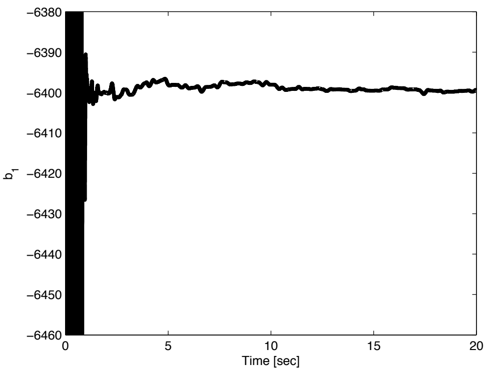
(b)



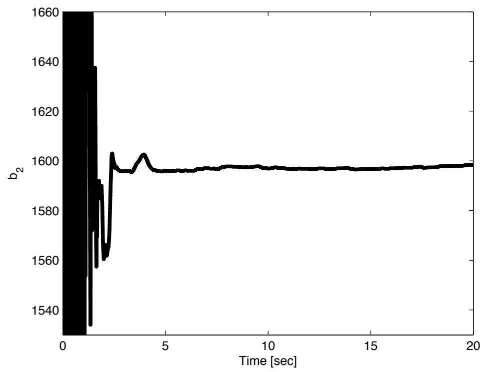
(c)



(d)



(e)



(f)

Fig. 4.15: The parameters estimated by the BC-NK estimator in noise conditions converge to the true values $b_1 = -6400$, $b_2 = 1600$, $a_1 = -5$, $a_2 = -408$, $a_3 = -416$ and $a_4 = -1600$.

4.3.1 Comparison with Hartley Modulating Function

The general formulation of the modulating function approaches was first developed by Shinbrot [95] in order to estimate the parameters of linear and non-linear systems. Further developments have been carried out and spawned several versions based on different modulating functions; they include the Fourier based functions either under a trigonometric form or under a complex exponential form: Spline-type functions, Hermite functions; Hartley-based functions. A very important advantage of using Fourier and Hartley-based modulating functions is that the system identification can be equivalently posed entirely in the frequency domain which makes it possible to use efficient DFT/FFT techniques. Both methods are well suited for digital implementation and have been included in the CONTSID toolbox ([30], [29] and [96]).

A function $\phi_{\mu,n}(t)$ is a modulating function of order n relative to a fixed time interval $[0, T]$, where μ is an index, if it is sufficiently smooth and possesses the following property for $l \in [0, n-1]$

$$\left[\frac{d^l \phi_{\mu,n}(t)}{dt^l}\right]_{t=0} = \left[\frac{d^l \phi_{\mu,n}(t)}{dt^l}\right]_{t=T} = 0$$

The modulating function and its first $(n-1)$ derivatives therefore vanishing at both end points of the observation from interval.
The Hartley Modulating function relies on the $cas(t)$ function defined by

$$cas(t) := \cos(t) + \sin(t)$$

Then is defined by:

$$\psi_{\mu,n} = \sum_{k=0}^{n} (-1)^k \binom{n}{k}$$

It is worth noting that the HMF method is real-valued; this presented an advantage since the input-output signals are real-valued [97].

A more detailed evaluation of the performances of the BC-NK algorithm can be carried out by looking at Tables (4.1),(4.2), (4.3), (4.4), (4.5) and (4.6) where the results obtained in a Monte Carlo study are presented (averaging over 200 runs). More precisely, the estimated parameters of the system under study (4.4) has been analysed (where $c_1 = -5$, $c_2 = -408$, $c_3 = -416$, $c_4 = -1600$, $k_1 = -6400$ and $k_2 = 1600$) using HMF and the proposed BC-NK technique, for increasing signal-to-noise ratio (SNR), where the SNR is defined as:

$$SNR = 10 \log_{10} \left(\frac{P_y}{P_\nu}\right) \tag{4.6}$$

P_ν represents the average power of the additive noise on the system output (*e.g.*the variance), while P_y denotes the average power of the noise-free output fluctuations. Note that in the table, each row provides the estimation of the unknown parameters for different levels of noise and sampling intervals Δt.
The table confirms in a more quantitative way the conclusions that were already arrived from the analysis led in this work. As can be seen, the developed approach provides accurate estimates of the parameters in every condition presented, of noise and sampling interval, showing a better behavior with smaller values of sampling interval, this is in agreement with the theoretical analysis that will be presented in Chapter 6. Comparised with HMF technique, the BC-NK shows good robustness at increasing SNR and good behaviors even when the sampling interval is greater (*i.e.*, when $\Delta t = 1 \cdot 10^{-3}$).

SNR	c_1	c_2	c_3	c_4	k_1	k_2
20	-4.685	-405.135	-413.231	-1568.656	-6250.399	1493.692
30	-4.919	-407.179	-414.543	-1597.616	-6384.814	1585.849
35	-4.957	-407.759	-415.413	-1599.539	-6394.814	1585.910

Table 4.1: Mean of the estimated parameters calculated using BC-NK approach with sampling interval $\Delta t = 1 \cdot 10^{-3}$ and different signal-to-noise ratio. The number of samples collected is $N_{samples} = 15011$

SNR	c_1	c_2	c_3	c_4	k_1	k_2
20	-5.079	-408.883	-450.807	-1666.169	-6389.525	848.298
30	-5.008	-408.113	-419.262	-1604.249	-6398.651	1519.295
35	-5.002	-408.036	-416.442	-1599.270	-6399.602	1582.447

Table 4.2: Mean of the estimated parameters calculated using HMF approach with sampling interval $\Delta t = 1 \cdot 10^{-3}$ and different signal-to-noise ratio. The number of samples collected is $N_{samples} = 15011$

SNR	c_1	c_2	c_3	c_4	k_1	k_2
20	-4.919	-406.793	-414.791	-1594.948	-6387.959	1579.771
30	-4.976	-407.453	-415.750	-1598.907	-6398.317	1594.118
35	-4.986	-407.871	-415.875	-1599.443	-6399.297	1596.752

Table 4.3: Mean of the estimated parameters calculated using BC-NK approach with sampling interval $\Delta t = 1 \cdot 10^{-4}$ and different signal-to-noise ratio. The number of samples collected is $N_{samples} = 100010$

SNR	c_1	c_2	c_3	c_4	k_1	k_2
20	-5.005	-408.180	-418.937	-1617.120	-6401.134	1440.808
30	-4.995	-408.001	-413.943	-1598.309	-6401.138	1608.519
35	-4.993	-407.967	-412.914	-1594.777	-6401.000	1641.983

Table 4.4: Mean of the estimated parameters calculated using HMF approach with sampling interval $\Delta t = 1 \cdot 10^{-4}$ and different signal-to-noise ratio. The number of samples collected is $N_{samples} = 100010$

SNR	c_1	c_2	c_3	c_4	k_1	k_2
20	-5.009	-407.515	-415.742	-1598.016	-6395.546	1589.910
30	-5.003	-407.892	-415.966	-1599.553	-6399.711	1596.998
35	-5.002	-407.944	-415.986	-1599.769	-6399.721	1598.333

Table 4.5: Mean of the estimated parameters calculated using BC-NK approach with sampling interval $\Delta t = 5 \cdot 10^{-5}$ and different signal-to-noise ratio. The number of samples collected is $N_{samples} = 200009$

SNR	c_1	c_2	c_3	c_4	k_1	k_2
20	-5.007	-408.154	-415.087	-1599.959	-6399.602	1568.693
30	-4.996	-407.999	-412.984	-1593.468	-6402.867	1643.441
35	-4.993	-407.967	-412.404	-1592.121	-6401.965	1660.997

Table 4.6: Mean of the estimated parameters calculated using HMF approach with sampling interval $\Delta t = 5 \cdot 10^{-5}$ and different signal-to-noise ratio. The number of samples collected is $N_{samples} = 200009$

4.3.2 Comparison with Refined Instrumental Variable Methodology

The Simplfied Refined Instrumental Variable method is a very powerful tool for the identification and estimation of continuous-time (SRIVC) Transfer Functions models. It was first suggested and implemented in 1980 (see [64]), while the full RIVC has been implemented recently (see [66] and [98]) and the reader should consult these publications for details.

Let's consider the benchmark example presented in (4.4) and let's deal a simulation experiment to compare BC-NK method and SRIVC method. We have considered two different scenarios: the first is obtained usign a $SNR = 35$ and a sampling interval of $\Delta t = 1 \cdot 10^{-3}$. The results of this numerical example has been reported in Figures (4.16),(4.17), (4.18), (4.19), (4.20) and (4.21), in which it have been highlighted the estimation of the parameters on a horizion of $t = 10$ seconds and its transient part of both methods. It is worth noting that the following resutls have been obtained computing the ergodic mean values, over a number of $N = 100$ runs. The second scenario, instead, is obtained with a $SNR = 30$ and a sampling interval of $\Delta t = 1 \cdot 10^{-4}$; a graphical representation of the results are shown in Figures (4.22),(4.23), (4.24), (4.25), (4.26) and (4.27).
From the analyzis of these behaviors, it is possible to deduce that the proposed technique shows, in the transient part of the estimates, very good results compared with SRIVC, while SRIVC have better or equal performances than BC-NK asymptotically. However, the transient behavior of the BC-NK is a relevant result, mainly due to the guaranteed internal stability conditions; besides, it is worth noting that unbounded transients could lead to saturation problems, hence have bounded transient is, for the proposed methodology, an important achievement.

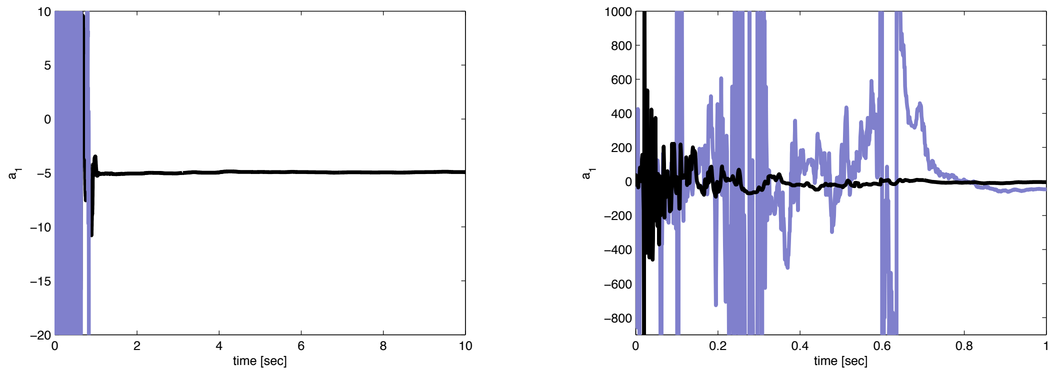


Fig. 4.16: Parameter $a_1 = -5$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

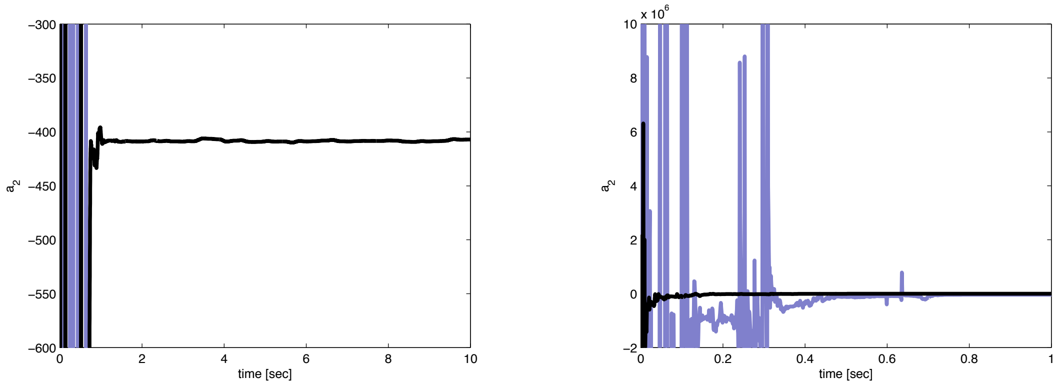


Fig. 4.17: Parameter $a_2 = -408$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

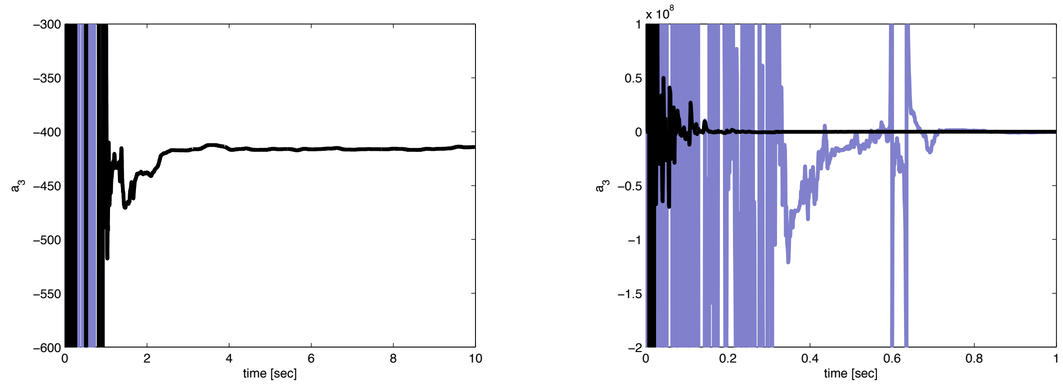


Fig. 4.18: Parameter $a_3 = -416$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

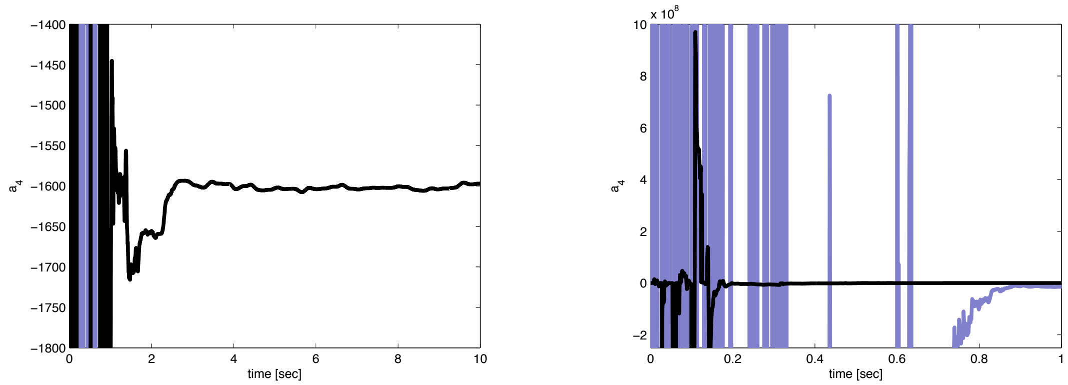


Fig. 4.19: Parameter $a_4 = -1600$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

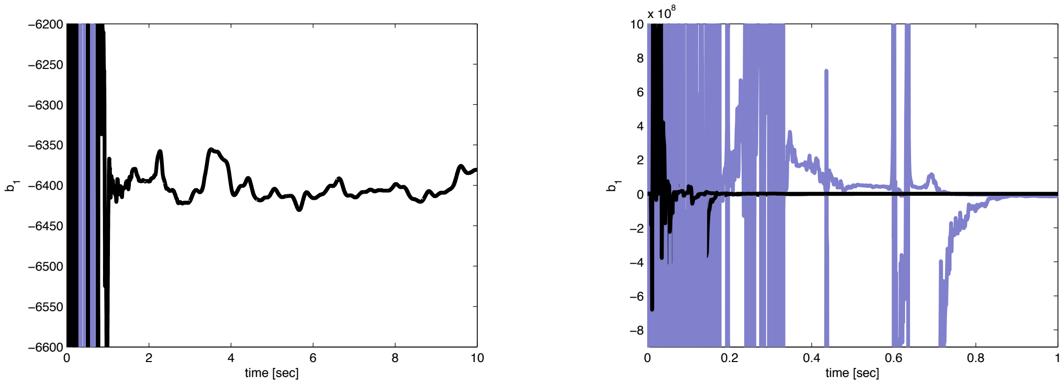


Fig. 4.20: Parameter $b_1 = -6400$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

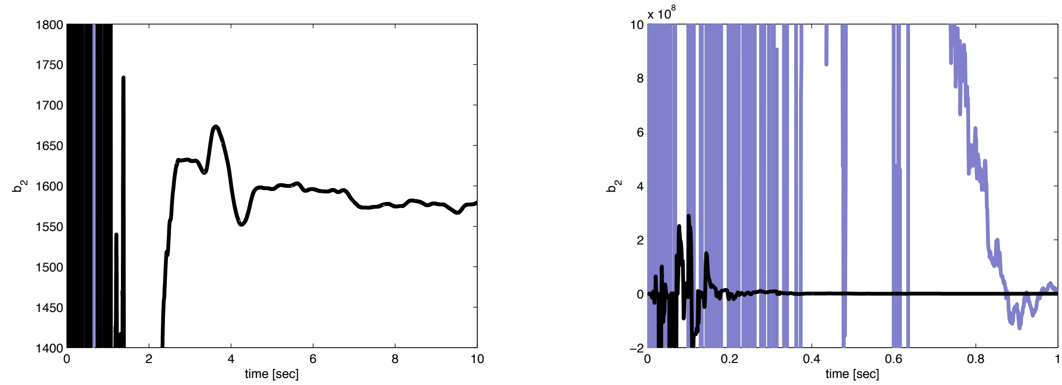


Fig. 4.21: Parameter $b_2 = 1600$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

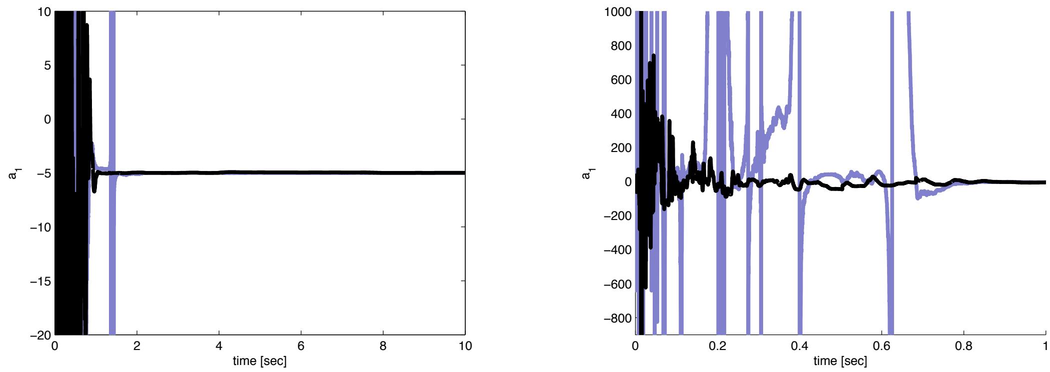


Fig. 4.22: Parameter $a_1 = -5$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

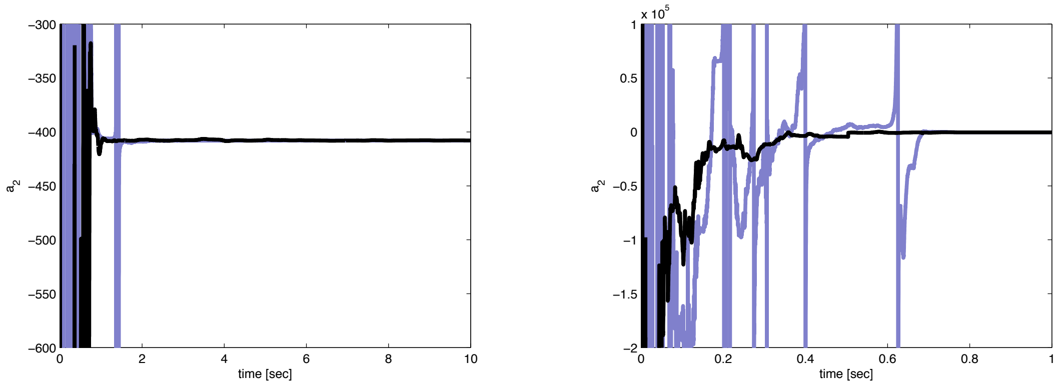


Fig. 4.23: Parameter $a_2 = -408$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

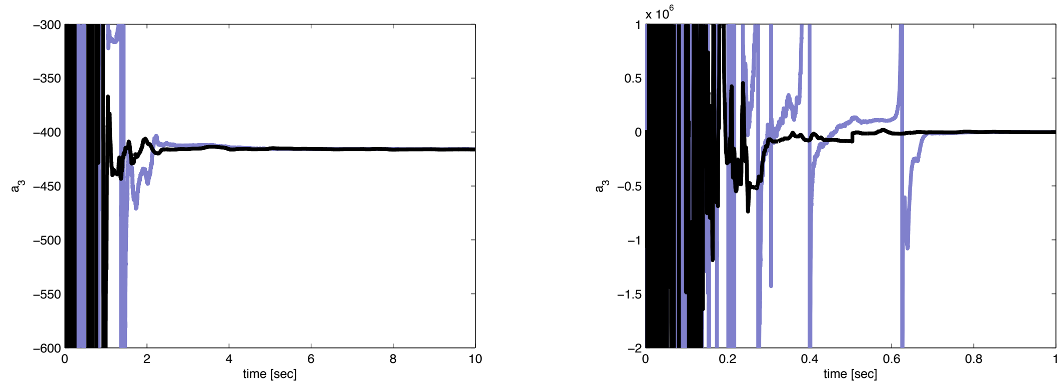


Fig. 4.24: Parameter $a_3 = -416$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

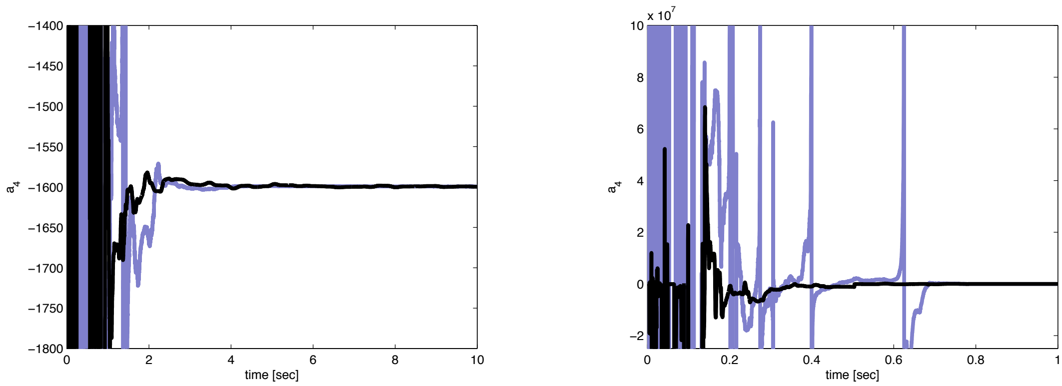


Fig. 4.25: Parameter $a_4 = -1600$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

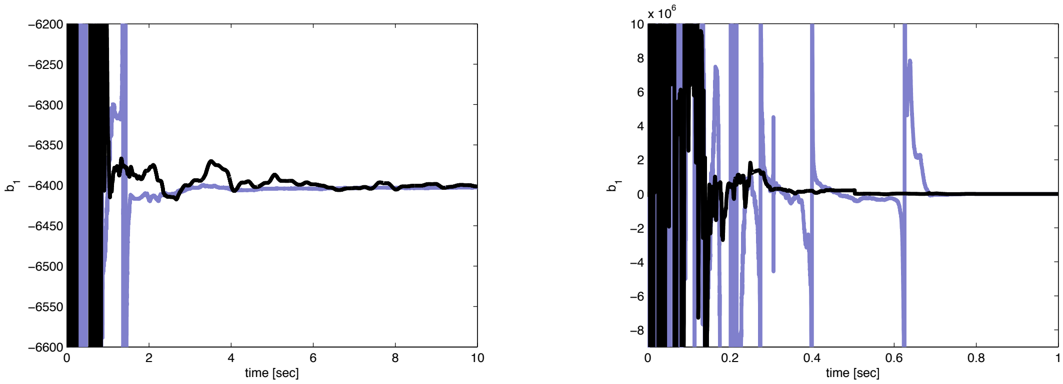


Fig. 4.26: Parameter $b_1 = -6400$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

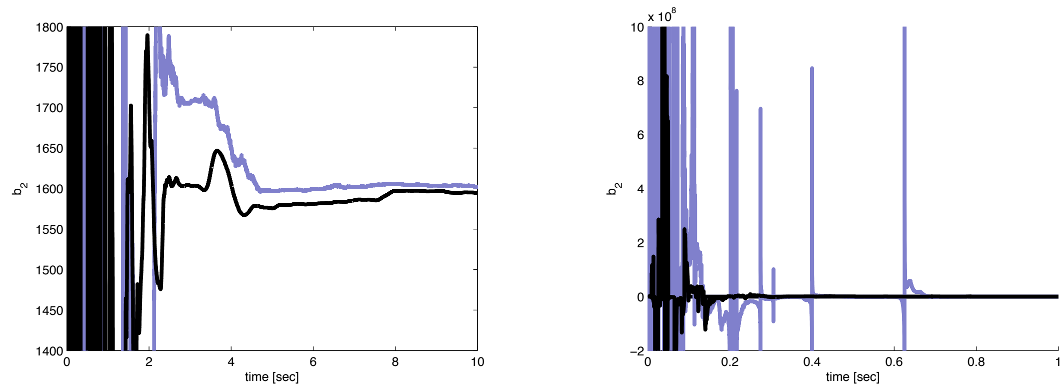


Fig. 4.27: Parameter $b_2 = 1600$ estimated by the SRIVC method (gray) and by the BC-NK estimator (black). The Figure on the right highlight the transient part of the simulation for both methods

Part II

Kernel-based State Estimation

Chapter 5

Non-asymptotic kernels for observers design

Let's consider again the Single-Input Single Output (SISO) CT system in input-output form (3.1), whence $m \in \mathbb{Z}_{>0}$, $n \in \mathbb{Z}_{>0}$, $m \leq n$ and $p \in \mathbb{Z}_{\geq 0}$. The values of the constant parameters $a_i \in \mathbb{R}$, with $i \in \{0, \ldots, n-1\}$ and $b_k \in \mathbb{R}$, with $k \in \{0, \ldots, m-1\}$ are assumed to be known. The only meaurable signals are $y(t)$ and $u(t)$, while their time-derivatives are not assumed to be available. Consider the following state-space realization (observer-canonical form) of (3.1):

$$\mathcal{S}_{u \to \mathbf{z} \to y} : \begin{cases} \mathbf{z}^{(1)}(t) = \mathbf{A}\,\mathbf{z}(t) + \mathbf{b}\,u(t), & t \in \mathbb{R}_{\geq 0} \\ y(t) = \mathbf{c}^\top \mathbf{z}(t), & \\ \mathbf{z}(0) = \mathbf{z}_0 & \end{cases} \tag{5.1}$$

where

$$\mathbf{z}(t) \triangleq [z_0(t)\ z_1(t)\ \ldots\ z_r(t)\ \ldots\ x_{n-1}(t)]^\top \in \mathbb{R}^n$$

is the system'state vector, while $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$ are given by

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} -a_{n-1} & 1 & 0 & \cdots & 0 \\ -a_{n-2} & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -a_1 & 0 & \cdots & 0 & 1 \\ -a_0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \\ \mathbf{b} &= \begin{bmatrix} 0 & \cdots & 0 & b_{m-1} & b_{m-2} & \cdots & b_1 & b_0 \end{bmatrix}^\top, \\ \mathbf{c} &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\top. \end{aligned} \tag{5.2}$$

For the sake of further discussion, it is worth to point out that the state-variables of the observer canonical realization can be expressed as a linear combination of the input-output derivatives:

$$z_r(t) = y^{(r)}(t) - \sum_{j=0}^{r-1} a_{n-r+j}\, y^{(j)}(t) - \sum_{j=0}^{r-1+m-n} b_{n-r+j}\, u^{(j)}(t)\,, \qquad r \in \{0, \ldots, n-1\}; \tag{5.3}$$

where we have used the convention $\sum_{j=0}^{k}\{\cdot\} = 0,\ \forall k < 0$. Our objective consists in providing a non-symptotic (fast) estimate of the state $\mathbf{z}(t)$ of system (5.1), by suitably processing the input and output signals $u(t)$ and $y(t)$, in such a way that the unknown value of the initial conditions $\mathbf{z}_0$ does not affect the transitory behavior of the estimator.

To this end, let $K \in \mathcal{HS}$ be a kernel satisfying the n-th order instantaneity condition:

$$K^{(j)}(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}, \forall j \in \{0,\dots,n-1\}. \tag{5.4}$$

This assumption implies that

$$\begin{aligned}
\left[V_K\, y^{(1)}\right](t) &= \int_0^t K(t,\tau)y^{(1)}(\tau)d\tau \\
&= y(t)K(t,t) - y(0)K(t,0) - \int_0^t K^{(1)}(t,\tau)y(\tau)d\tau \\
&= y(t)K(t,t) - \int_0^t K^{(1)}(t,\tau)y(\tau)d\tau
\end{aligned}$$

and therefore

$$\left[V_{K^{(1)}}\, y\right](t) = y(t)K(t,t) - \left[V_K\, y^{(1)}\right](t) \tag{5.5}$$

Note that (5.5) holds for any instantaneous kernel of order n-th and for any function y admitting a first generalized derivative. Changing the integrand y with $y^{(i)}$ and the kernel K with $K^{(j)}$, for some $j \in \{1,\dots,n-1\}$, under the assumption $K^{(j)}(t,0) = 0$, $\forall t \in \mathbb{R}_{\geq 0}$, we have that also the following integral equation holds

$$\left[V_{K_h^{(j+1)}} y^{(i)}\right](t) = y^{(i)}(t)K_h^{(j)}(t,t) - \left[V_{K_h^{(j)}} y^{(i+1)}\right](t) \tag{5.6}$$

Thanks to (5.6), we can arrange the following system of n integral equations:

$$\begin{cases}
\left[V_{K^{(n)}} y\right](t) = y(t)K^{(n-1)}(t,t) - \left[V_{K^{(n-1)}} y^{(1)}\right](t) \\
\left[V_{K^{(n-1)}} y^{(1)}\right](t) = y^{(1)}(t)K^{(n-2)}(t,t) - \left[V_{K^{(n-2)}} y^{(2)}\right](t) \\
\qquad \vdots \\
\left[V_{K^{(n-i)}} y^{(i)}\right](t) = y^{(i)}(t)K^{(n-i-1)}(t,t) - \left[V_{K^{(n-i-1)}} y^{(i+1)}\right](t) \\
\qquad \vdots \\
\left[V_{K^{(1)}} y^{(n-1)}\right](t) = y^{(n-1)}(t)K(t,t) - \left[V_K\, y^{(n)}\right](t)
\end{cases} \tag{5.7}$$

Introducing the dynamic constraint in the last equation we get the useful relation

$$\begin{aligned}\left[V_{K^{(1)}}y^{(n-1)}\right](t) &= y^{(n-1)}(t)K(t,t) - \left[V_K\, y^{(n)}\right](t)\\ &= y^{(n-1)}(t)K(t,t) - \left[V_K\left(\sum_{i=0}^{n-1}a_i y^{(i)}(t) + \sum_{k=0}^{m-1}b_k u^{(k)}(t)\right)\right](t)\\ &= y^{(n-1)}(t)K(t,t) - \sum_{i=0}^{n-1}a_i\left[V_K y^{(i)}\right](t) - \sum_{k=0}^{m-1}b_k\left[V_K u^{(k)}\right](t)\end{aligned} \tag{5.8}$$

The following result will be exploited in the sequel.

Lemma 5.0.1 (Technical) *Let $K \in \mathcal{HS}$ verify the n-th order instantaneity condition $K^{(j)}(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}$, $\forall j \in \{0,\ldots,n-1\}$, and let $y(\cdot) \in \mathcal{L}^2$ admit a n-th generalized derivative, then for $i \in \{1,\ldots,n-1\}$:*

$$\left[V_{K^{(n-i)}}y^{(i)}\right](t) = \sum_{j=0}^{i-1}(-1)^{i-1+j}y^{(j)}(t)K^{(n-j-1)}(t,t) + (-1)^i[V_{K^{(n)}}y](t), \ . \tag{5.9}$$

□

Proof 4 *The system of equations (5.7) can be rearranged as:*

$$\begin{cases}\left[V_{K^{(n-1)}}y^{(1)}\right](t) &= y(t)K^{(n-1)}(t,t) - \left[V_{K^{(n)}}\,y\right](t)\\ \left[V_{K^{(n-2)}}y^{(2)}\right](t) &= y^{(1)}(t)K_h^{(n-2)}(t,t) - \left[V_{K^{(n-1)}}y^{(1)}\right](t)\\ &\vdots\\ \left[V_{K^{(n-i)}}y^{(i)}\right](t) &= y^{(i-1)}(t)K^{(n-i)}(t,t) - \left[V_{K^{(n-i+1)}}y^{(i-1)}\right](t)\\ \left[V_{K^{(n-i-1)}}y^{(i+1)}\right](t) &= y^{(i)}(t)K^{(n-i-1)}(t,t) - \left[V_{K^{(n-i)}}y^{(i)}\right](t)\\ &\vdots\\ \left[V_K\ y^{(n)}\right](t) &= y^{(n-1)}(t)K(t,t) - \left[V_{K^{(1)}}y^{(n-1)}\right](t)\end{cases} \tag{5.10}$$

By substituting the expression $y(t)K^{(n-1)}(t,t) - [V_K\, y](t)$ in place of $[V_{K_h^{(n-1)}}y^{(1)}](t)$ in the right side of the second equation and proceeding by forward substitution into the successive equations, the statement of the lemma can be trivially obtained.

Consider now (5.9) with $i = n-1$:

$$\left[V_{K^{(1)}}y^{(n-1)}\right](t) = \sum_{j=0}^{n-2}(-1)^{n-2+j}y^{(j)}(t)K^{(n-j-1)}(t,t) + (-1)^{n-1}[V_{K^{(n)}}y](t) \tag{5.11}$$

Changing the kernel $K^{(n-i)}$ in (5.9) with K we also obtain

$$\left[V_K\ y^{(i)}\right](t) = \sum_{j=0}^{i-1}(-1)^{i-1+j}y^{(j)}(t)K^{(i-j-1)}(t,t) + (-1)^i[V_{K^{(i)}}y](t), \;\; i \in \{1,\dots,n-1\} \tag{5.12}$$

Changing the integrand $y^{(i)}$ in (5.12) with $u^{(k)}$:

$$\left[V_K\ u^{(k)}\right](t) = \sum_{j=0}^{k-1}(-1)^{k-1+j}u^{(j)}(t)K^{(k-j-1)}(t,t) + (-1)^k[V_{K^{(k)}}u](t), \;\; i \in \{1,\dots,m-1\} \tag{5.13}$$

Substituting (5.11),(5.12) and (5.13) in (5.8) we get an equation that relates $y^{(n-1)}$ with lower-order derivatives $y^{(1)}(t),\dots,y^{(n-2)}(t)$, $u^{(1)}(t),\dots,u^{(m-1)}(t)$ and with the non-anticipative signals $y(t)$, $u(t)$ and $[V_K\ y](t)$, $[V_K\ u](t)$:

$$\begin{aligned}(-1)^{n-1}[V_{K^{(n)}}y](t) = {} & y^{(n-1)}(t)K(t,t) - \sum_{j=0}^{n-2}(-1)^{n-2+j}y^{(j)}(t)K^{(n-j-1)}(t,t) \\ & - \sum_{i=0}^{n-1} a_i \left(\sum_{j=0}^{i-1}(-1)^{i-1+j}y^{(j)}(t)K^{[i-j-1]}(t,t) + (-1)^i[V_{K^{(i)}}y](t)\right) \\ & - \sum_{k=0}^{m-1} b_k \left(\sum_{j=0}^{k-1}(-1)^{k-1+j}u^{(j)}(t)K^{(k-j-1)}(t,t) + (-1)^k[V_{K^{(k)}}u](t)\right)\end{aligned} \tag{5.14}$$

which can be rearranged as

$$\begin{aligned}(-1)^{n-1}[V_{K^{(n)}}y](t) + \sum_{i=0}^{n-1} a_i(-1)^i[V_{K^{(i)}}y](t) + \sum_{k=0}^{m-1} b_k(-1)^k[V_{K^{(k)}}u](t) = \\ = \; y^{(n-1)}(t)K(t,t) - \sum_{j=0}^{n-2}(-1)^{n+j}y^{(j)}(t)K^{(n-j-1)}(t,t) \\ + \sum_{i=0}^{n-1} a_i \sum_{j=0}^{i-1}(-1)^{i+j}y^{(j)}(t)K^{(i-j-1)}(t,t) \\ + \sum_{k=0}^{m-1} b_k \sum_{j=0}^{k-1}(-1)^{k+j}u^{(j)}(t)K^{(k-j-1)}(t,t)\end{aligned} \tag{5.15}$$

By first introducing the substitution $k = i$ on the right side of equation (5.15) and by exploiting the technical result (8.3), reported in the Appendix, we can change the order of summation in the nested indexed sums, obtaining

$$
\begin{aligned}
&(-1)^{n-1}[V_{K^{(n)}}y](t)+\sum_{i=0}^{n-1}a_i(-1)^i[V_{K^{(i)}}y](t)+\sum_{k=0}^{m-1}b_k(-1)^k[V_{K^{(k)}}u](t)=\\
&\quad=y^{(n-1)}(t)K(t,t)+\sum_{j=0}^{n-2}y^{(j)}(t)\Big((-1)^{n+j+1}K^{(n-j-1)}(t,t)+\sum_{i=j+1}^{n-1}a_i(-1)^{i+j}K^{(i-j-1)}(t,t)\Big)\\
&\quad+\sum_{j=0}^{m-2}u^{(j)}(t)\sum_{i=j+1}^{m-1}b_i(-1)^{i+j}K^{(i-j-1)}(t,t)=y^{(n-1)}(t)K(t,t)\\
&\quad+\sum_{j=0}^{n-2}y^{(j)}(t)\Big((-1)^{n+j+1}K^{(n-j-1)}(t,t)++\sum_{p=j+2}^{n}a_{n-p+1+j}(-1)^{n-p+1}K^{(n-p)}(t,t)\Big)+\\
&\quad+\sum_{j=0}^{m-2}u^{(j)}(t)\sum_{p=j+2+n-m}^{n}b_{n-p+1+j}(-1)^{n-p+1}K^{(n-p)}(t,t)
\end{aligned}
\tag{5.16}
$$

where we have also posed $p=j+n-i+1$.
Now, in view of (8.4) and (8.5) (reported in the Appendix), we can rewrite (5.16) as

$$
\begin{aligned}
&(-1)^{n-1}[V_{K^{(n)}}y](t)+\sum_{i=0}^{n-1}a_i(-1)^i[V_{K^{(i)}}y](t)+\sum_{k=0}^{m-1}b_k(-1)^k[V_{K^{(k)}}u](t)=\\
&=\sum_{p=1}^{n}K^{(n-p)}(t,t)\,(-1)^{n-p}y^{(p-1)}(t)+\sum_{p=2}^{n}K^{(n-p)}(t,t)\sum_{j=0}^{p-2}(-1)^{n+1-p}\,a_{n-p+j+1}\,y^{(j)}(t)+\\
&+\sum_{p=2+n-m}^{n}K^{(n-p)}(t,t)\sum_{j=0}^{p-2+m-n}(-1)^{n+1-p}\,b_{n-p+j+1}\,u^{(j)}(t)=\\
&=\sum_{r=0}^{n-1}K^{(n-r-1)}(t,t)\,(-1)^{n-r-1}y^{(r)}(t)-\sum_{r=1}^{n-1}K^{(n-r-1)}(t,t)(-1)^{n-r-1}\sum_{j=0}^{r-1}a_{n-r+j}\,y^{(j)}(t)-\\
&-\sum_{r=1+n-m}^{n-1}K^{(n-r-1)}(t,t)(-1)^{n-r-1}\sum_{j=0}^{r-1+m-n}b_{n-r+j}\,u^{(j)}(t)=\\
&=K^{(n-1)}(t,t)\,(-1)^{n-1}y(t)+\sum_{r=1}^{n-m}K^{(n-r-1)}(t,t)\,(-1)^{n-r-1}\Big(y^{(r)}(t)-\sum_{j=0}^{r-1}a_{n-r+j}\,y^{(j)}(t)\Big)+\\
&+\sum_{r=1+n-m}^{n-1}K^{(n-r-1)}(t,t)(-1)^{n-r-1}\Big(y^{(r)}(t)-\sum_{j=0}^{r-1}a_{n-r+j}\,y^{(j)}(t)-\sum_{j=0}^{r-1+m-n}b_{n-r+j}\,u^{(j)}(t)\Big)
\end{aligned}
\tag{5.17}
$$

where we have finally posed $r=p-1$. Considering that all the terms in the left side of (5.17) can be obtained by causal filtering the signals y and u, it is convenient define the auxiliary signal

$$\nu(t) \triangleq (-1)^{n-1}[V_{K^{(n)}}y](t) + \sum_{i=0}^{n-1} a_i(-1)^i[V_{K^{(i)}}y](t) + \sum_{k=0}^{m-1} b_k(-1)^k[V_{K^{(k)}}\,u](t) \tag{5.18}$$

and then express (5.17) in compact form as

$$\sum_{p=0}^{n-1} \gamma_r(t)\, z_r(t) = \nu(t)\,, \tag{5.19}$$

where $\gamma_0(t), \ldots, \gamma_r(t), \ldots, \gamma_{n-1}(t)$ are known functions of time depending on the particular kernel function chosen for the implementation of the estimator:

$$\gamma_r(t) = K^{(n-r-1)}(t,t)(-1)^{n-r-1}\,,\ \ r \in \{0, \ldots, n-1\} \tag{5.20}$$

while $z_0(t), \ldots, z_r(t), \ldots, z_{n-1}(t)$ are unknown signals (depending on unavaliable time-derivatives of u and y) corresponding to the state variables of a state-space realization of the system, and depending, in turn, on the unkown derivatives of y and u:

$$z_r(t) = \begin{cases} y(t), & r = 0; \\ y^{(r)}(t) - \sum_{j=0}^{r-1} a_{n-r+j}\, y^{(j)}(t)\,, & 1 \le r < \min(1+n-m, n-1); \\ y^{(r)}(t) - \sum_{j=0}^{r-1} a_{n-r+j}\, y^{(j)}(t) - \sum_{j=0}^{r-1+m-n} b_{n-r+j}\, u^{(j)}(t)\,, & 1+n-m \le r \le n-1; \end{cases} \tag{5.21}$$

An algebraic system having a number of equations of the kind (5.19) equal to the number of unknowns (i.e., the instantaneous values of the state variables $z_0(t), \ldots, z_{n-1}(t)$) can be arranged by using n different kernel functions $K_1, K_2, \ldots, K_n$ to enforce new independent constraints. Clearly, to ensure the invertibility of the system with respect to the unknown state vector for any $t > 0$, we will introduce further constraints on the structure of the kernel functions.

5.1 Non-asymptotic state estimation with instantaneous internally stable kernels

Now, the basic requirement $K_h \in \mathcal{HS}$ ensures that, for any bounded $y(\cdot)$, the transformed signals $[V_{K^{(i)}}y](\cdot)$ have not finite-time escape. Now, for the estimator to be implementable as an internally stable filter, we also need to ensure the asymptotic boundedness of all the signals involved in the estimation process. Modulating functions can be designed to achieve this task (e.g., $\tau^n e^{-\omega\tau}, ...$), but these result in integrals with time-varying weighting patterns that asymptotically decay to 0. Thus the internal stability comes at the price of a practical freezing of the filter as time proceeds from the activation instant. We seek for a non-fading kernel keeping the estimator alive undefinitely.

Now, we will propose a family of admissible $\mathcal{HS}$ kernels, realizable by means of a finite-dimensional asymptotically stable state-space implementation ([4]), fulfilling the n-th order instantaneity condition, being $n \in \mathbb{Z}_{>0}$ the order of the system $\mathcal{S}_{u\to y}$ under concern.

Let us consider the following parametrized kernel

$$K_{\omega:\varpi:\mu}(t,\tau) \;\; = \;\; e^{-\omega\,(t-\tau)}\left(1-e^{-\varpi\tau}\right)^{\mu}$$

with $\mu \in \mathbb{R}_{>0}$ and $\omega, \varpi \in \mathbb{R}_{>0}$ arbitrary constant parameters. For any particular choice of ω and ϖ, an admissible kernel function, verifying the n-th order instantaneity conditions , can be obtained by setting $\mu \geq n$. Indeed the derivative kernel functions write:

$$\begin{aligned} K^{(i)}_{\omega:\varpi:\mu}(t,\tau) &= \frac{\overline{\partial}^i}{\overline{\partial}\tau^i} K_{\omega:\varpi:\mu}(t,\tau) \\ &= \frac{\partial^i}{\partial\tau^i}\left(e^{-\omega t}e^{\omega\tau}(1-e^{-\varpi\tau})^{\mu}\right) \\ &= e^{-\omega t}\,\frac{\mathrm{d}^i}{\mathrm{d}\tau^i}\Big[e^{\omega\tau}(1-e^{-\varpi\tau})^{\mu}\Big], \;\; i \in \{0,\ldots,n\}. \end{aligned} \tag{5.22}$$

for which it holds that

$$K^{(i)}_{\omega:\varpi:\mu}(t,t) \;\; = \;\; e^{-\omega t}\,\frac{\mathrm{d}^i}{\mathrm{d}\tau^i}\Big[e^{\omega\tau}(1-e^{-\varpi\tau})^{\mu}\Big]\;\Bigg|_{\tau=t}$$

is nontrivial for all $0 \leq i \leq n$. Moreover

$$K^{(i)}_{\omega:\varpi:\mu}(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}, \forall\, j \in \{0\ldots n-1\}.$$

Note that each derivative kernel $K^{(i)}_{\omega:\varpi:\mu}(t,\tau)$, although separable, is not of convolution type, which implies that the opeerator $V_{K^{(i)}_{\omega:\varpi:\mu}}$ does not admit a linear time-invariant dynamical system realization. Nonetheless, we will show that, for any bounded signal $x(t)$ the transformed signal $[V_{K^{(i)}_{\omega:\varpi:\mu}} x](t)$ can be obtained as the output of a nonlinear stable scalar dynamical system, for any $i \in \{0, \dots, n\}$.

In order to prove this statement, let us specialize (2.8) to the kernel $K^{(i)}_{\omega:\varpi:\mu}(t,\tau)$ defined in (5.22)

$$\begin{cases} \xi^{(1)}(t) = \begin{cases} K^{(i)}_{\omega:\varpi:\mu}(t,t)x(t) + \int\limits_0^t \left(\dfrac{\partial}{\partial t} K^{(i)}_{\omega:\varpi:\mu}\right) x(\tau) d\tau, & t \in \mathbb{R}_{\geq 0}; \\ 0, & t = 0; \end{cases} \\ \xi(0) = \xi_0 = 0; \end{cases}$$
$$\left[V_{K^{(i)}_{\omega:\varpi:\mu}} x\right](t) = \xi(t), \ \ \forall t \in \mathbb{R}_{\geq 0}. \tag{5.23}$$

Considering that

$$\frac{\partial}{\partial t} K^{(i)}_{\omega:\varpi:\mu}(t,\tau) = -\omega K^{(i)}_{\omega:\varpi:\mu}(t,\tau),$$

observing that $K^{(i)}_{\omega:\varpi:\mu}(0,0)$ is bounded, and assuming that $x(t)$ is bounded too, then a state-space realization of the operator induced by the kernel under concern is given by:

$$\begin{aligned} \xi^{(1)}(t) &= K^{(i)}_{\omega:\varpi:\mu}(t,t)x(t) + \int\limits_0^t \left(\frac{\partial}{\partial t} K^{(i)}_{\omega:\varpi:\mu}\right) x(\tau) d\tau \\ &= K^{(i)}_{\omega:\varpi:\mu}(t,t)x(t) - \omega \int\limits_0^t K^{(i)}_{\omega:\varpi:\mu}(t,\tau)x(\tau) d\tau \\ &= K^{(i)}_{\omega:\varpi:\mu}(t,t)x(t) - \omega \left[V_{K^{(i)}_{\omega:\varpi:\mu}} x\right](t) \\ &= K^{(i)}_{\omega:\varpi:\mu}(t,t)x(t) - \omega\, \xi(t), \qquad \forall t \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{5.24}$$

Being $K^{(i)}_{\omega:\varpi:\mu}(t,t)$ bounded , it holds that the scalar dynamical system realization of the Volterra operators induced by the proposed kernels is internally stable, and ISS (Input-to-state stable) w.r.t. $x(t)$.

Now, let us exploit the proposed class of kernel functions to solve the state estimation problem. For the sake of enhancing the clarity of the presentation, given a given system of order n, let us reduce the kernel parametric structure by fixing $\varpi \in \mathbb{R}_{>0}$ arbitrarily and by posing $\mu = n$ (in this way we assign to μ the minimum value of exponent under which the n-th order instantaneity condition is preserved):

$$K_{\omega}(t,\tau) = e^{-\omega t}\left[e^{\omega\tau}\left(1-e^{-\varpi\tau}\right)^{n}\right], \tag{5.25}$$

where we have removed the parameters ϖ and n, that will remain fixed and shared by all the kernels considered in the sequel. Consider n kernels of the kind (5.25) with different ω parameters, that is:
$K_{\omega_0}(t,\tau),\ldots,K_{\omega_{n-1}}(t,\tau)$ with

$$\omega_0,\ldots,\omega_{n-1} \in \mathbb{R}^n_{>0} : \ \omega_i \neq \omega_j, \forall i,j \in \{0\ldots,n-1\}^2, i \neq j. \tag{5.26}$$

Now, incolumnating n equations of the kind (5.19), built by the kernels K_{ω_h}, with $h \in \{0,\ldots,n-1\}$, we get the following time-verying system linear equations (alebraic for any specific $t \in \mathbb{R}_{\geq 0}$)

$$\mathbf{\Gamma}(t)\mathbf{z}(t) \ = \ \boldsymbol{\nu}(t) \tag{5.27}$$

where $\mathbf{z}(t) \triangleq [z_0(t),\ldots,z_{n-1}(t)]^T \in \mathbb{R}^n$ is the unkwnown vector of states to be estimated, $\boldsymbol{\nu}(t) \triangleq [\nu_0(t),\ldots,\nu_{n-1}(t)]^T \in \mathbb{R}^n$ is a vector of known signals (obtainable by non-anticipative operators) and $\mathbf{K}(t) \in \mathbb{R}^{n\times n}$ is a square time-varying matrix defined as:

$$\begin{aligned}
\mathbf{\Gamma}(t) \ &\triangleq \ \left|\begin{array}{ccccc}
\gamma_{0,0} & \cdots & \gamma_{0,r} & \cdots & \gamma_{0,n-1} \\
\gamma_{1,0} & \cdots & \gamma_{1,r} & \cdots & \gamma_{1,n-1} \\
\vdots & & \vdots & & \vdots \\
\gamma_{h,0} & \cdots & \gamma_{h,r} & \cdots & \gamma_{h,n-1} \\
\vdots & & \vdots & & \vdots \\
\gamma_{n-1,0} & \cdots & \gamma_{n-1,r} & \cdots & \gamma_{n-1,n-1}
\end{array}\right| \\
&= \left|\begin{array}{ccccc}
(-1)^{n-1}K^{(n-1)}_{\omega_0}(t,t) & \cdots & (-1)^{n-r-1}K^{(n-r-1)}_{\omega_0}(t,t) & \cdots & K_{\omega_0}(t,t) \\
(-1)^{n-1}K^{(n-1)}_{\omega_1}(t,t) & \cdots & (-1)^{n-r-1}K^{(n-r-1)}_{\omega_1}(t,t) & \cdots & K_{\omega_1}(t,t) \\
\vdots & & \vdots & & \vdots \\
(-1)^{n-1}K^{(n-1)}_{\omega_h}(t,t) & \cdots & (-1)^{n-r-1}K^{(n-r-1)}_{\omega_h}(t,t) & \cdots & K_{\omega_h}(t,t) \\
\vdots & & \vdots & & \vdots \\
(-1)^{n-1}K^{(n-1)}_{\omega_{n-1}}(t,t) & \cdots & (-1)^{n-r-1}K^{(n-r-1)}_{\omega_1}(t,t) & \cdots & K_{\omega_{n-1}}(t,t)
\end{array}\right|.
\end{aligned} \tag{5.28}$$

Remark 5.1.1 (Direct feedthrough and invertibility of (5.27)) *It is worth to point out that the algebraic system (5.27) can be solved with respect to $\mathbf{z}(t)$ if the matrix $\mathbf{\Gamma}(t)$ is invertible. It holds that $\mathbf{\Gamma}(t)$ is invertible for any $t > 0$, being the rows of $\mathbf{\Gamma}(t)$ mutually independent thanks to the devised* *kernel functions. This feature is implied by the particular class of kernels chosen for the implementation of the estimator,* i.e., *the BF-NK type.*

Compared with the Bivariate (strict) Causal Non-asymptotic Kernel (BC-NK) (Section 3) to implement the non-instantaneous parameter-estimator, for which $C(t,t) = 0$, $\forall t \in \mathbb{R}_{\geq 0}$, *it holds instead that for the BF-NK (5.22)* $F^{(i)}(t,t) > 0$, $\forall t > 0$. *This property will imply that the state-space realization of the operator* V_F *induced by the kernel, applied to the derivative of a signal,* $[V_F u^{(1)}](t)$, *can be written in a form tha has a direct feedtrough from* $u(t)$*;* i.e., $[V_F u^{(1)}](t) = F(t,t)u(t) - [V_{F^{(1)}} u](t)$. *All the elements of* $\mathbf{\Gamma}(t)$ *will therefore be positive for any* $t > 0$.

Moreover, $\mathbf{\Gamma}(t)$ *is guaranteed to be full rank (invertible) for any* $t > 0$ *because of condition (5.26).* □

Our focus is now directed to the computation $\boldsymbol{\nu}(t)$. In view of (5.33), each elements $\nu_h(t), h \in \{0, \dots, n-1\}$ is given by a weighted sum of the signals u, y and their transformations, obtainable by applying integral operators to u and y.

In view of the result (5.24), concerning the state-space realization of the integral operators induced by the proposed kernel functions, it is possible to conclude that the non-asymptotic state estimator can be implemented as an internally stable dynamic system.

Indeed, defining the internal state vector

$$\boldsymbol{\xi}(t) = [\boldsymbol{\xi}_{\omega_0}(t), \boldsymbol{\xi}_{\omega_1}(t), \dots, \boldsymbol{\xi}_{\omega_{n-1}}(t)]^T \in \mathbb{R}^{(n+1+m)n},$$

with

$$\boldsymbol{\xi}_{\omega_i}(t) = \begin{bmatrix} [V_{K_{\omega_i}} y](t), [V_{K^{(1)}_{\omega_i}} y](t), \dots, [V_{K^{(n-1)}_{\omega_i}} y](t), [V_{K^{(n)}_{\omega_i}} y](t) \\ [V_{K_{\omega_i}} u](t), [V_{K^{(1)}_{\omega_i}} u](t), \dots, [V_{K^{(m-1)}_{\omega_i}} u](t) \end{bmatrix}^T \in \mathbb{R}^{n+1+m},$$

where $i \in \{0, \dots, n-1\}$; it is now possible to assign the followig structure to $\boldsymbol{\nu}(t)$:

$$\boldsymbol{\nu}(t) = \mathbf{H}\boldsymbol{\xi}(t), \tag{5.29}$$

where, in view of (5.18), the matrices in (5.29) are given by

$$\mathbf{H} = \begin{vmatrix} \mathbf{h}^T & 0 & \cdots & 0 \\ 0 & \mathbf{h}^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{h}^T \end{vmatrix} \in \mathbb{R}^{n \times [(n+1+m)n]},$$

with

$$\mathbf{h}^T = \Big| a_0,\ -a_1,\ \cdots,\ a_i(-1)^i,\ \cdots,\ a_{n-1}(-1)^{n-1},\ (-1)^{n-1},$$
$$b_0,\ -b_1,\ \ldots,\ b_k(-1)^k,\ \cdots,\ b_{m-1}(-1)^{m-1} \Big| \quad \in \mathbb{R}^{1\times(n+1+m)},$$

In view of (5.24), the time evolution of the state vetor $\boldsymbol{\xi}(t)$ is described by the following system of ordinary differential equations

$$\begin{cases} \boldsymbol{\xi}^{(1)}(t) &= \mathbf{G}\boldsymbol{\xi}(t) + \mathbf{E}(t)y(t) + \mathbf{F}(t)u(t), \quad t \in \mathbb{R}_{\geq 0}, \\ \boldsymbol{\xi}(0) &= \mathbf{0}, \end{cases} \tag{5.30}$$

Notably, the matrix $\mathbf{G}$ is diagonal, time-invariant and Hurwitz:

$$\mathbf{G} = \begin{vmatrix} \mathbf{G}_{\omega_0} & 0 & \cdots & 0 \\ 0 & \mathbf{G}_{\omega_1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_{\omega_{n-1}} \end{vmatrix} \in \mathbb{R}^{[(n+1+m)n]\times[(n+1+m)n]}$$

$$\mathbf{G}_{\omega_i} = \begin{vmatrix} -\omega_i & 0 & \cdots & 0 \\ 0 & -\omega_i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\omega_i \end{vmatrix} \in \mathbb{R}^{(n+1+m)\times(n+1+m)}$$

while the time-varying input matrices can be expressed as

$$\mathbf{E}(t) = \begin{vmatrix} \mathbf{E}_{\omega_0}(t) \\ \mathbf{E}_{\omega_1}(t) \\ \vdots \\ \mathbf{E}_{\omega_{n-1}}(t) \end{vmatrix} \in \mathbb{R}^{(n+1+m)n}, \quad \mathbf{F}(t) = \begin{vmatrix} \mathbf{F}_{\omega_0}(t) \\ \mathbf{F}_{\omega_1}(t) \\ \vdots \\ \mathbf{F}_{\omega_{n-1}}(t) \end{vmatrix} \in \mathbb{R}^{(n+1+m)n}$$

where

$$\mathbf{E}_{\omega_i}(t) = \begin{vmatrix} K_{\omega_i}(t,t) \\ K_{\omega_i}^{(1)}(t,t) \\ \vdots \\ K_{\omega_i}^{(n-1)}(t,t) \\ K_{\omega_i}^{(n)}(t,t) \\ 0 \\ 0 \\ \vdots \\ 0 \end{vmatrix} \in \mathbb{R}^{n+1+m}, \quad \mathbf{F}_{\omega_i}(t) = \begin{vmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ K_{\omega_i}(t,t) \\ K_{\omega_i}^{(1)}(t,t) \\ \vdots \\ K_{\omega_i}^{(m-1)}(t,t) \end{vmatrix} \in \mathbb{R}^{n+1+m}.$$

Finally, by (5.27), (5.29) and (5.30), the state estimator takes the following form:

$$\begin{cases} \boldsymbol{\xi}^{(1)}(t) & = & \mathbf{G}\boldsymbol{\xi}(t) + \mathbf{E}(t)y(t) + \mathbf{F}(t)u(t), \;\; t \in \mathbb{R}_{\geq 0}, \\ \boldsymbol{\xi}(0) & = & \mathbf{0}, \end{cases}$$
$$\mathbf{z}(t) = \Big(\mathbf{\Gamma}(t)\Big)^{-1} \mathbf{H}\boldsymbol{\xi}(t), \;\; t \in \mathbb{R}_{>0}. \tag{5.31}$$

that is, a stable linear time-varying dynamical system.

By analyzing the time-trend of the elements of $\mathbf{\Gamma}(t)$ we can conclude that the outcome of the estimator (that is, the estimated state vector $\mathbf{z}(t)$) is defined almost everywhere. Indeed, for any $t > 0$, the matrix $\mathbf{\Gamma}(t)$ is invertible, while, at the activation time instant $t = 0$, $\mathbf{\Gamma}(0)$ is singular. Conversely, for any $t > 0$, the outcome of the estimator is a vector of continuous-in-time signals that, in nominal conditions, are exact non-delayed estimates of the states of the observed system.

Remark 5.1.2 (Separation Principle) *An important observation is that the $\boldsymbol{\xi}$ dynamics, driven by the signals $u(t), y(t)$, depends only on the particular kernels chosen for the implementation. Indeed the constant Hurwitz matrix $\mathbf{G}$ and the time-varying matrices $\mathbf{E}(t)$ and $\mathbf{F}(t)$ can be expressed in terms of the sole kernel functions. The parameters of the system $\mathcal{S}_{u\to y}$ (that is, $a_0, \ldots, a_{n-1}$ and $b_0, \ldots, b_{m-1}$) only affect the output map $\mathbf{H}$ and the time-varying output maps $\mathbf{J}(t)$ and $\mathbf{L}(t)$. In conlusion, the internal stability properties of the estimator can be structurally enforced by the choice of the kernel functions and can be made independent from the parameters of the system under observation.*

Note that, in the usual observer design methods (see [99], [73]), based on the synthesis of a correction gain from the measured output, the design procedure requires the knowledge of the system's parameters to compute the correction gain, in order to assign the poles of the estimator (Luenberger) or to optimize an optimality criterion (Kalman) guaranteing the stability of the estimator's dynamics. Thus, compared to usual filter design methods, in our setting the internal dynamics of the observer can be designed neglecting at all the parameters of the system under concern. Most important, the parameters of the system can be used only when available, while the information on the past evolution of the system can be retained and stored in the internal states of the filter.

In conclusion: we do not use high gain correction feedback neither differentiators, instead we have obtained a non-anticipative state-space (finite-dimensional) implementation with internal stability, with assigned observer dynamics regardless of the dynamics of the observed system, and , most important, instantaneous (non-asymtoptic) estimates for any $t > 0$.

5.2 Reducing the complexity of the estimator

Now, with the aim of reducing the complexity of the estimator, we will introduce some modification to the basic formulation given in Section 5.1. Assuming that u e y are not affected by disturbances, being u and y directly measurable, and therefore already available, it is possible to remove those vaiables from the vector of estimates, and reduce the dimensions of the problem.

To this end, let us collect all the measurable and computable quantities (i.e., the signals $u(t)$,$y(t)$ and their transformations produced by implementable non-anticipative operators) on the left side of (5.16):

$$
\begin{aligned}
&(-1)^{n-1}[V_K^{(n)}y](t)+\sum_{i=0}^{n-1}a_i(-1)^i[V_{K(i)}y](t)+\sum_{k=0}^{m-1}b_k(-1)^i[V_{K(k)}u](t)\\
&+\left(K^{(n-1)}(t,t)\,(-1)^n+\sum_{p=2}^{n}K^{(n-p)}(t,t)(-1)^{n-p}\,a_{n-p+1}\right)y(t)\\
&+\sum_{p=2+n-m}^{n}\quad K^{(n-p)}(t,t)\,(-1)^{n-p}\,b_{n-p+1}\,u(t)\quad=\\
&=\;K_h^{(n-2)}(t,t)\,(-1)^{n-2}y^{(1)}(t)+\;\sum_{p=3}^{2+m-n}K_h^{(n-p)}(t,t)\,(-1)^{n-p}\left(y^{(p-1)}(t)-\sum_{j=1}^{p-2}a_{n-p+j+1}\,y^{(j)}(t)\right)\\
&+\sum_{p=3+n-m}^{n}K_h^{(n-p)}(t,t)\,(-1)^{n-p}\left(y^{(p-1)}(t)-\sum_{j=1}^{p-2+n-m}b_{n-p+j+1}\,u^{(j)}(t)-\sum_{j=1}^{p-2}a_{n-p+j+1}\,y^{(j)}(t)\right)\\
&=\;K_h^{(n-2)}(t,t)\,(-1)^{n-2}y^{(1)}(t)+\;\sum_{r=2}^{1+m-n}K_h^{(n-r-1)}(t,t)\,(-1)^{n-r-1}\left(y^{(r)}(t)-\sum_{j=1}^{r-1}a_{n-r+j}\,y^{(j)}(t)\right)\\
&+\sum_{p=2+n-m}^{n-1}K_h^{(n-r-1)}(t,t)\,(-1)^{n-r-1}\left(y^{(r)}(t)-\sum_{j=1}^{r-1+n-m}b_{n-r+j}\,u^{(j)}(t)-\sum_{j=1}^{r-1}a_{n-r+j}\,y^{(j)}(t)\right)
\end{aligned}
\tag{5.32}
$$

where we have posed $r=p-1$. Considering that all the terms in the left side of (5.32) can be obtained by non-anticipative filtering the signals y

and u, we can define the auxiliary signal

$$\begin{aligned}\nu_h(t) \triangleq\, & (-1)^{n-1}[V_K^{(n)} y](t) + \sum_{i=0}^{n-1} a_i (-1)^i [V_{K^{(i)}} y](t) + \sum_{i=0}^{m-1} b_i (-1)^i [V_{K^{(i)}} u](t) \\ & + \left(K^{(n-1)}(t,t)\,(-1)^n + \sum_{i=1}^{n-1} K^{(n-i-1)}(t,t)(-1)^{n-i-1}\, a_{n-i} \right) y(t) \\ & + \sum_{i=1+n-m}^{n-1} K^{(n-i-1)}(t,t)\,(-1)^{n-i-1}\, b_{n-i}\, u(t) \ ,\end{aligned} \tag{5.33}$$

and then express (5.32) in compact form as

$$\sum_{r=1}^{n-1} \gamma_r(t)\, z_r(t) = \nu(t)\,, \tag{5.34}$$

where $\gamma_r(t)$ are known functions defined in (5.20), while $z_1(t), \ldots, z_r(t), \ldots, z_{n-1}(t)$ are unknown signals (depending on unavaliable time-derivatives of u and y) corresponding to the state variables of a state-space realization of the system, and depending, in turn, on the unkown derivatives of y and u:

$$z_r(t) = \begin{cases} y^{(1)}(t), & r = 1; \\ y^{(r)}(t) - \sum_{j=1}^{r-1} a_{n-r+j} y^{(j)}(t)\,, & \\ \qquad 2 \le r < \min(1+n-m, n-1); & \\ y^{(r)}(t) - \sum_{j=1}^{r-1} a_{n-r+j} y^{(j)}(t) - \sum_{j=1}^{r-1+m-n} b_{n-r+j} u^{(j)}(t)\,, & \\ \qquad 2+n-m \le r \le n-1; & \end{cases} \tag{5.35}$$

An algebraic system having a number of equations of the kind (5.34) equal to the number of unknowns (i.e., the instantaneous values of the state variables $z_1(t), \ldots, z_{n-1}(t)$) can be arranged by using $(n-1)$ different kernel functions $K_1, K_2, \ldots, K_{n-1}$ to enforce new independent constraints.

Now, let us exploit the proposed class of kernel functions to solve the state estimation problem. Now, let us consider $n-1$ kernels of the kind (5.25) with different ω parameters, that is: $K_{\omega_1}(t,\tau), \ldots, K_{\omega_{n-1}}(t,\tau)$ with

$$\omega_1, \ldots, \omega_{n-1} \in \mathbb{R}_{>0}^{n-1} : \ \omega_i \neq \omega_j, \forall i,j \in \{1, \ldots, n-1\}^2, i \neq j.$$

Now, incolumnating $n-1$ equations of the kind (5.34), built by the kernels K_{ω_h}, with $h \in \{1, \ldots, n-1\}$, we get a system of linear equations analogous to (5.27):

$$\mathbf{\Gamma}(t)\mathbf{z}(t) = \boldsymbol{\nu}(t)$$

where the dimension of the vector of unknowns is reduced by one unit with respect to (5.27):

$$\mathbf{z}(t) \triangleq [z_1(t), \ldots, z_{n-1}(t)]^T \in \mathbb{R}^{n-1}$$

as long as

$$\boldsymbol{\nu}(t) \triangleq [\nu_1(t), \ldots, \nu_{n-1}(t)]^T \in \mathbb{R}^{n-1}$$

, while $\mathbf{\Gamma}(t) \in \mathbb{R}^{n-1 \times n-1}$ can be written as:

$$\begin{aligned}
\mathbf{\Gamma}(t) &\triangleq \begin{vmatrix} \gamma_{1,1} & \cdots & \gamma_{1,r} & \cdots & \gamma_{1,n-1} \\ \gamma_{2,1} & \cdots & \gamma_{2,r} & \cdots & \gamma_{2,n-1} \\ \vdots & & \vdots & & \vdots \\ \gamma_{h,1} & \cdots & \gamma_{h,r} & \cdots & \gamma_{h,n-1} \\ \vdots & & \vdots & & \vdots \\ \gamma_{n-1,1} & \cdots & \gamma_{n-1,r} & \cdots & \gamma_{n-1,n-1} \end{vmatrix} \\
&= \begin{vmatrix} (-1)^{n-1}K_{\omega_1}^{(n-2)}(t,t) & \cdots & (-1)^{n-r-1}K_{\omega_1}^{(n-r-1)}(t,t) & \cdots & K_{\omega_1}(t,t) \\ (-1)^{n-1}K_{\omega_2}^{(n-2)}(t,t) & \cdots & (-1)^{n-r-1}K_{\omega_1}^{(n-r-1)}(t,t) & \cdots & K_{\omega_2}(t,t) \\ \vdots & & \vdots & & \vdots \\ (-1)^{n-1}K_{\omega_h}^{(n-2)}(t,t) & \cdots & (-1)^{n-r-1}K_{\omega_h}^{(n-r-1)}(t,t) & \cdots & K_{\omega_h}(t,t) \\ \vdots & & \vdots & & \vdots \\ (-1)^{n-1}K_{\omega_{n-1}}^{(n-2)}(t,t) & \cdots & (-1)^{n-r-1}K_{\omega_1}^{(n-r-1)}(t,t) & \cdots & K_{\omega_{n-1}}(t,t) \end{vmatrix}.
\end{aligned} \tag{5.36}$$

Following the same lines of Section 5.1 it is possible to show that the non-asymptotic state estimator can be implemented as an internally stable dynamic system. Indeed, defining the internal state vector

$$\boldsymbol{\xi}(t) = [\boldsymbol{\xi}_{\omega_1}(t), \boldsymbol{\xi}_{\omega_2}(t), \ldots, \boldsymbol{\xi}_{\omega_{n-1}}(t)]^T \in \mathbb{R}^{(n+1+m)(n-1)},$$

with

$$\boldsymbol{\xi}_{\omega_i}(t) = \begin{matrix}\Big[[V_{K_{\omega_i}}y](t), [V_{K_{\omega_i}^{(1)}}y](t), \ldots, [V_{K_{\omega_i}^{(n-1)}}y](t), [V_{K_{\omega_i}^{(n)}}y](t) \\ [V_{K_{\omega_i}}u](t), [V_{K_{\omega_i}^{(1)}}u](t), \ldots, [V_{K_{\omega_i}^{(m-1)}}u](t)\Big]^T\end{matrix} \in \mathbb{R}^{n+1+m}.$$

it is possible to assign the followig structure to $\boldsymbol{\nu}(t)$:

$$\boldsymbol{\nu}(t) = \mathbf{H}\boldsymbol{\xi}(t) + \mathbf{L}(t)u(t) + \mathbf{J}(t)y(t), \tag{5.37}$$

where, in view of (5.33), the matrices in (5.29) are given by

$$\mathbf{H} = \begin{vmatrix} \mathbf{h}^T & 0 & \cdots & 0 \\ 0 & \mathbf{h}^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{h}^T \end{vmatrix} \in \mathbb{R}^{(n-1)\times[(n+1+m)(n-1)]},$$

with

$$\mathbf{h}^T = \Big| a_0,\ -a_1,\ \cdots,\ a_i(-1)^i,\ \cdots,\ a_{n-1}(-1)^{n-1},\ (-1)^{n-1},$$
$$b_0,\ -b_1,\ \ldots,\ b_k(-1)^k,\ \cdots,\ b_{m-1}(-1)^{m-1} \Big| \quad \in \mathbb{R}^{1\times(n+1+m)},$$

and

$$\mathbf{J}(t) = \left| \begin{array}{c} K_{\omega_1}^{(n-1)}(t,t)(-1)^n + \sum_{j=1}^{n-1} K_{\omega_1}^{(n-j-1)}(t,t)(-1)^{n-j-1}a_{n-j} \\ K_{\omega_2}^{(n-1)}(t,t)(-1)^n + \sum_{j=1}^{n-1} K_{\omega_2}^{(n-j-1)}(t,t)(-1)^{n-j-1}a_{n-j} \\ \vdots \\ K_{\omega_{n-1}}^{(n-1)}(t,t)(-1)^n + \sum_{j=1}^{n-1} K_{\omega_{n-1}}^{(n-j-1)}(t,t)(-1)^{n-j-1}a_{n-j} \end{array} \right| \in \mathbb{R}^{n-1},$$

$$\mathbf{L}(t) = \left| \begin{array}{c} \sum_{j=1+n-m}^{n-1} K_{\omega_1}^{(n-j-1)}(t,t)(-1)^{n-j-1}b_{n-j} \\ \sum_{j=1+n-m}^{n-1} K_{\omega_2}^{(n-j-1)}(t,t)(-1)^{n-j-1}b_{n-j} \\ \vdots \\ \sum_{j=1+n-m}^{n-1} K_{\omega_{n-1}}^{(n-j-1)}(t,t)(-1)^{n-j-1}b_{n-j} \end{array} \right| \in \mathbb{R}^{n-1}.$$

In view of (5.24), the time evolution of the state vetor $\boldsymbol{\xi}(t)$ is described by the a system of ordinary differential equations having the form of (5.30). In this case the, matrix $\mathbf{G}$ takes the form:

$$\mathbf{G} = \left| \begin{array}{cccc} \mathbf{G}_{\omega_1} & 0 & \cdots & 0 \\ 0 & \mathbf{G}_{\omega_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_{\omega_{n-1}} \end{array} \right| \in \mathbb{R}^{[(n+1+m)(n-1)]\times[(n+1+m)(n-1)]}$$

$$\mathbf{G}_{\omega_i} = \left| \begin{array}{cccc} -\omega_i & 0 & \cdots & 0 \\ 0 & -\omega_i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\omega_i \end{array} \right| \in \mathbb{R}^{(n+1+m)\times(n+1+m)}$$

while the time-varying input matrices can be expressed as

$$\mathbf{E}(t) = \left| \begin{array}{c} \mathbf{E}_{\omega_1}(t) \\ \mathbf{E}_{\omega_2}(t) \\ \vdots \\ \mathbf{E}_{\omega_{n-1}}(t) \end{array} \right| \in \mathbb{R}^{(n+1+m)(n-1)}, \quad \mathbf{F}(t) = \left| \begin{array}{c} \mathbf{F}_{\omega_1}(t) \\ \mathbf{F}_{\omega_2}(t) \\ \vdots \\ \mathbf{F}_{\omega_{n-1}}(t) \end{array} \right| \in \mathbb{R}^{(n+1+m)(n-1)}$$

where

$$
\mathbf{E}_{\omega_i}(t) = \begin{vmatrix} K_{\omega_i}(t,t) \\ K^{(1)}_{\omega_i}(t,t) \\ \vdots \\ K^{(n-1)}_{\omega_i}(t,t) \\ K^{(n)}_{\omega_i}(t,t) \\ 0 \\ 0 \\ \vdots \\ 0 \end{vmatrix} \in \mathbb{R}^{n+1+m}, \quad \mathbf{F}_{\omega_i}(t) = \begin{vmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ K_{\omega_i}(t,t) \\ K^{(1)}_{\omega_i}(t,t) \\ \vdots \\ K^{(m-1)}_{\omega_i}(t,t) \end{vmatrix} \in \mathbb{R}^{n+1+m}.
$$

Finally, by (5.27), (5.37) and (5.30), the state estimator takes the following form:

$$
\begin{aligned}
&\begin{cases} \boldsymbol{\xi}^{(1)}(t) &= \mathbf{G}\boldsymbol{\xi}(t) + \mathbf{E}(t)y(t) + \mathbf{F}(t)u(t), \quad t \in \mathbb{R}_{\geq 0}, \\ \boldsymbol{\xi}(0) &= \mathbf{0}, \end{cases} \\
&\mathbf{z}(t) = \Big(\boldsymbol{\Gamma}(t)\Big)^{-1}\Big(\mathbf{H}\boldsymbol{\xi}(t) + \mathbf{J}(t)y(t) + \mathbf{L}(t)u(t)\Big), \quad t \in \mathbb{R}_{>0}.
\end{aligned}
\tag{5.38}
$$

that is, a stable linear time-varying dynamical system with a reduced state dimension with respect to the full-dimensional filter (5.31) presented in Section 5.1.

5.3 Conluding remarks

In this Chapter the Bivariate Feedthrough Non-Asymptotic Kernels have been introduced; the results obtained follow the ones derived in Chapter 3; however, this Chapter provided a different design for the kernel function in order to obtain an "instantaneous" estimation for the states of a dynamic system. In Chapter 7, some simulation results will be presented in order to show the effectiveness of the proposed state estimation architecture.

Chapter 6

Asymptotic analysis of the kernel-based continuous-time model identification algorithm

In this part the problem of analysing the asymptotic properties of the proposed identification algorithm is considered [5]. To this purpose, the so-called hybrid framework of continuous-time model identification (see [44]) is assumed, *i.e.*, the system generating the data is assumed to be a continuous-time, linear time-invariant system, while the noise model is defined in discrete-time for the sake of simplicity.

6.1 Bias analysis

Consider the continuous-time linear time-invariant input-output system

$$x^{(n)}(t) = \sum_{i=0}^{n-1} a_i x^{(i)}(t) + \sum_{i=0}^{m-1} b_i u^{(i)}(t), \tag{6.1}$$

corresponding to (3.1) with $p = 1$, and introduce the following assumptions:

Assumption 1 *the system* (6.1) *is asymptotically stable.*

Assumption 2 *The input u is a quasi stationary, piece-wise constant, deterministic sequence.*

Consider now a dataset consisting of K input-output measurements associated with the sampling instants $t_k = t_0 + kT$, $k = 0, \ldots, K-1$ (uniform sampling is assumed, for the sake of simplicity), defined as follows

$$y(k) = x(t_k) + e(k)$$

$$u(k) = u(t_k),$$

where $e(k)$ represents (output) measurement noise, are available. Then, the following assumptions hold:

Assumption 3 *the process e is a stationary zero mean white process noise with second moments*

$$\mathbb{E}\left[e(k)e^T(j)\right] = R_e\delta_{ij}$$

where δ_{ij} is the Kronecker delta.

Assumption 4 *The input u is uncorrelated with the noise e.*

Assumption 5 *Instantaneous sampling, in the sense of [100], is assumed,* i.e., *sampling is assumed to be fast with respect to the dynamics of interest.*

Assumption 6 *In view of an asymptotic analysis, $G = H(\tau)$ is chosen as kernel of the augmentation operator V_G.*

To deal with deterministic and stochastic signals in a compact manner, the following operator is defined

$$\bar{\mathbb{E}}[\cdot] = \lim_{N\to\infty} \frac{1}{K}\sum_{t=1}^{K} \mathbb{E}[\cdot]$$

where $\mathbb{E}[\cdot]$ is the expectation operator. For two signals $a(t)$ and $b(t)$, the cross covariance matrix will be denoted as $\mathbf{R}_{ab} = \bar{\mathbb{E}}\left[a(t)b^T(t)\right]$.

Then, the identification algorithm proposed in Section 3 aims at estimating the parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} a_0 & a_1 & \dots & a_{n-1} & b_0 & b_1 & \dots & b_{m-1} \end{bmatrix}^T = \begin{bmatrix} \boldsymbol{\theta}_y^T & \boldsymbol{\theta}_u^T \end{bmatrix}$$

on the basis of the available data by solving the linear regression problem

$$r_{y,n+p} = \mathbf{z}^T\boldsymbol{\theta} \tag{6.2}$$

where

$$\mathbf{z} = \begin{bmatrix} r_{y,p} & \dots & r_{y,p+n-1} & r_{u,p} & \dots & r_{u,p+m-1} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_y & \mathbf{z}_u \end{bmatrix} \tag{6.3}$$

and $r_{y,i}$,$r_{u,i}$ are given by the outputs of a suitably discretised version of the filter bank in equation (3.18), to be defined in the following. Note that the state space representation of the filter bank in (3.18) can be broken down to a set of $n+1$ filters for the output y

$$\dot{\boldsymbol{\xi}}_{y,i}(t) = \mathbf{G}\boldsymbol{\xi}_{y,i}(t) + \mathbf{E}_i(t)y(t) \tag{6.4}$$

$$r_{y,i}(t) = \mathbf{H}\boldsymbol{\xi}_{y,i}(t), \tag{6.5}$$

$i = p, \ldots, p+n$ and m filters for the input u

$$\begin{aligned}\dot{\boldsymbol{\xi}}_{u,i}(t) &= \mathbf{G}\boldsymbol{\xi}_{u,i}(t) + \mathbf{E}_i(t)u(t) && (6.6)\\ r_{u,i}(t) &= \mathbf{H}\boldsymbol{\xi}_{u,i}(t), && (6.7)\end{aligned}$$

$i = p, \ldots, p+m-1$, where

$$\mathbf{G} = \operatorname{diag}\begin{bmatrix}-\omega & \ldots & -\omega(N+1)\end{bmatrix}$$

and

$$\mathbf{H} = \begin{bmatrix}1 & \ldots & 1\end{bmatrix}.$$

For the purpose of an asymptotic analysis, the time-varying matrices $E_i(t)$ have to be replaced with their asymptotic values, so in the following the constant matrices

$$\bar{\mathbf{E}}_i = \lim_{t\to\infty} \mathbf{E}_i(t) \tag{6.8}$$

will be considered in the definition of the filters in equations (6.5) and (6.7). Note that in view of the definitions of the kernels giving rise to the time-varying vectors $\mathbf{E}_i(t)$, the limits in (6.8) are well defined.

In the following we will denote with $F_i(s)$ the transfer function associated with the state space quadruple $\big(\mathbf{G}, \bar{\mathbf{E}}_i, \mathbf{H}, 0\big)$. For the purpose of the following analysis it is interesting to point out and exploit the particular structure of the filters $F_i(s)$: indeed, as $\mathbf{G}$ is diagonal and the definition of the output matrix $\mathbf{H}$ implies that the output of each filter is simply given by the sum of its states, one can conclude that $F_i(s)$ can be written as

$$F_i(s) = \sum_{j=1}^{N+1} F_{ij}(s), \tag{6.9}$$

with

$$F_{ij}(s) = \frac{\bar{E}_{ij}}{z - G_j}, \quad j = 1, \ldots, N+1 \tag{6.10}$$

where $G_j = -j\omega$, $j = 1, \ldots, N+1$ and where $\bar{E}_{ij}$ denotes the jth element of vector $\bar{\mathbf{E}}_i$.

Finally, as the hybrid framework of continuous-time identification has been assumed, for the sake of implementation a discretised version of the above defined continuous-time filters has to be derived. By using, *e.g.*, the backward Euler transformation, the discrete-time counterparts of the $F_{ij}(s)$ filters can be derived as

$$F_{ij}(z) = F_{ij}(s)|_{s=(z-1)/Tz} = \frac{\bar{E}_{d,ij}}{s - G_{d,j}}, \quad j = 1, \ldots, N+1$$

where

$$\bar{E}_{d,ij} = \frac{\bar{E}_{ij}T}{1 - G_j T}, \tag{6.11}$$

$$G_{d,j} = \frac{1}{1 - G_j T}, \tag{6.12}$$

so that

$$F_i(z) = \sum_{j=1}^{N+1} F_{ij}(z).$$

Therefore, in discrete-time, the variables appearing in the regression (6.2) and in (6.3) can be defined as

$$r_{y,i}(k) = F_i(z)y(k), i = p, \dots, n+p$$

$$r_{u,i}(k) = F_i(z)u(k), i = p, \dots, m-1+p.$$

Finally, in the definition of $r_{y,i}$ it is convenient to highlight the deterministic part, resulting from the filtering of $x(k)$ and the stochastic part, resulting from the filtering of $e(k)$, as follows:

$$r_{y,i}(k) = r_{y,i}(k) + e_i(k), i = p, \dots, n+p$$

where

$$r_{x,i}(k) = F_i(z)x(k), i = p, \dots, n+p$$

and

$$e_i(k) = F_i(z)e(k), i = p, \dots, n+p.$$

Similarly, $\mathbf{z}_y$ in (6.3) can be expressed as $\mathbf{z}_y = \mathbf{z}_x + \mathbf{z}_e$, with obvious definitions of $\mathbf{z}_x$ and $\mathbf{z}_e$.

On the basis of the above definitions, the aim is to establish an expression for the bias of the estimate of $\boldsymbol{\theta}$ computed by solving the discrete-time regression.

Now we want to lead the bias computation, to this end it is convenient to express the linear regression (6.2) as

$$r_{y,n+p} = \mathbf{z}^T\boldsymbol{\theta} = \begin{bmatrix}\mathbf{z}_x^T + \mathbf{z}_e^T & \mathbf{z}_u^T\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}_y \\ \boldsymbol{\theta}_u\end{bmatrix} = \begin{bmatrix}\mathbf{z}_x^T & \mathbf{z}_u^T\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}_y \\ \boldsymbol{\theta}_u\end{bmatrix} + \mathbf{z}_e^T\boldsymbol{\theta}_y. \tag{6.13}$$

Left-multiplying by $\mathbf{z}$ equation (6.13) and letting $\mathbf{z}_{x,u} = \begin{bmatrix}\mathbf{z}_x & \mathbf{z}_u\end{bmatrix}$, one gets

$$\begin{aligned}\mathbf{z}_{x,u}r_{y,n+p} = \mathbf{z}_{x,u}\mathbf{z}_{x,u}^T\boldsymbol{\theta} - \begin{bmatrix}\mathbf{z}_e \\ 0\end{bmatrix} r_{x,n+p} - \begin{bmatrix}\mathbf{z}_e \\ 0\end{bmatrix} e_{n+p} + \\ + \begin{bmatrix}\mathbf{z}_e \\ 0\end{bmatrix}\begin{bmatrix}\mathbf{z}_x^T & \mathbf{z}_u^T\end{bmatrix} + \begin{bmatrix}\mathbf{z}_x\mathbf{z}_e^T + \mathbf{z}_e\mathbf{z}_e^T & 0 \\ \mathbf{z}_u\mathbf{z}_e^T & 0\end{bmatrix}\boldsymbol{\theta}.\end{aligned} \tag{6.14}$$

Letting now

$$\mathbf{R}_{\mathbf{z}_{xu} r_{x,n+p}} = \bar{\mathbb{E}}\left[\mathbf{z}_{xu} r_{x,n+p}\right],$$
$$\mathbf{R}_{\mathbf{z}_{xu}} = \bar{\mathbb{E}}\left[\mathbf{z}_{xu}\mathbf{z}_{xu}^T\right],$$

and applying the $\bar{\mathbb{E}}[\cdot]$ operator to both sides of (6.14) one gets

$$\mathbf{R}_{\mathbf{z}_{x,u} r_{y,n+p}} = \mathbf{R}_{\mathbf{z}_{x,u}}\boldsymbol{\theta} - \mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix} e_{n+p}} + \mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix}}\boldsymbol{\theta},$$

which, regrouping terms, can be written as

$$(\mathbf{R}_{\mathbf{z}_{x,u} r_{y,n+p}} - \mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix} e_{n+p}}) = (\mathbf{R}_{\mathbf{z}_{x,u}} + \mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix}})\boldsymbol{\theta}. \tag{6.15}$$

Therefore it is clear from (6.15) that in the absence of measurement noise the regression reduces to

$$\mathbf{R}_{\mathbf{z}_{x,u} r_{y,n+p}} = \mathbf{R}_{\mathbf{z}_{x,u}}\boldsymbol{\theta}, \tag{6.16}$$

which leads to an unbiased estimate. When noise is taken into account, the resulting estimate is necessarily affected by bias if a least squares solution of the linear regression problem in (6.15) is considered. While this is a known fact in the continuous-time identification literature, which has led to the development of sophisticated instrumental variable algorithms for bias elimination (see, *e.g.*, [44] and the references therein), it is interesting to pursue the above analysis further, exploiting the above derived expressions for the discrete-time counterparts of the filters (6.5) and (6.7).

More precisely, letting $\boldsymbol{\theta}^o$ the true value of the unknown parameter vector (corresponding to the solution of the noise-free regression (6.16)) and denoting with $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^o$ the bias in the computed noisy estimate, it is easy to see from (6.15) and (6.16) that

$$\mathbb{E}\left[\Delta\boldsymbol{\theta}\right] = [\mathbf{R}_{\mathbf{z}_{x,u}} + \mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix}}]^{-1}[-\mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix} e_{n+p}} - \mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix}}\boldsymbol{\theta}^o]. \tag{6.17}$$

The noise-dependent covariance functions $\mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix}}$ and $\mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix} e_{n+p}}$ in (6.17) can be further analysed by noting that

$$\mathbf{R}_{\begin{bmatrix}\mathbf{z}_e\\0\end{bmatrix}} = \begin{bmatrix}\mathbb{E}\left[\mathbf{z}_e\mathbf{z}_e^T\right] & 0\\ 0 & 0\end{bmatrix},$$

where, in turn,

$$\mathbb{E}\left[\mathbf{z}_e\mathbf{z}_e^T\right] = \mathbb{E}\begin{bmatrix} e_p^2 & e_p e_{p+1} & \dots & e_p e_{p+n-1} \\ \vdots & \vdots & & \vdots \\ e_p e_{p+n-1} & e_{p+1} e_{p+n-1} & \dots & e_{p+n-1}^2 \end{bmatrix},$$

and

$$\mathbf{R}_{\begin{bmatrix}\mathbf{z}_e \\ 0\end{bmatrix} e_{n+p}} = \mathbb{E}\begin{bmatrix} e_p e_{n+p} \\ e_{p+1} e_{n+p} \\ \vdots \\ e_{p+n-1} e_{n+p} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

So, to evaluate the bias on the estimated parameters, one has to compute

$$R_{e_p e_q} = \mathbb{E}\left[e_p e_q\right], \quad q = 0, \dots, n.$$

To this purpose, note that by definition

$$\mathbb{E}\left[e_p e_q\right] = \mathbb{E}\left[(F_p(z)e)(F_q(z)e)\right],$$

which, in view of (6.9), can be written as

$$\mathbb{E}\left[(F_p(z)e)(F_q(z)e)\right] = \mathbb{E}\left[(\sum_{j=1}^{N+1} F_{p,j}(z)e)(\sum_{l=1}^{N+1} F_{q,l}(z)e)\right] = \mathbb{E}\left[(\sum_{j=1}^{N+1} e_{p,j})(\sum_{l=1}^{N+1} e_{q,l})\right],$$

where (see (6.10))

$$e_{p,j}(k) = F_{p,j}(z)e(k) = \frac{\bar{E}_{dp,j}}{1 - G_{dj}z^{-1}} e(k),$$

or, equivalently,

$$e_{p,j}(k) = G_{dj} e_{p,j}(k-1) + \bar{E}_{dp,j} e(k-1),$$

and similarly for $e_{q,l}$.

It follows that the covariance

$$R_{e_{p,j} e_{q,l}} = \mathbb{E}\left[e_{p,j}(k) e_{q,l}(k)\right]$$

corresponds to the covariance between two first order AR processes forced by the same white noise input, so that

$$R_{e_{p,j} e_{q,l}} = \frac{\bar{E}_{dp,j}\bar{E}_{dq,l}}{1 - G_{dj}G_{dl}} R_e, \tag{6.18}$$

and, in turn,

$$R_{e_p e_q} = \sum_{j=1}^{N+1} \sum_{l=1}^{N+1} \frac{\bar{E}_{dp,j} \bar{E}_{dq,l}}{1 - G_{dj} G_{dl}} R_e,$$

which can be used in (6.17) to quantify the bias in the estimate of $\boldsymbol{\theta}^o$.

The above results on the asymptotic expression of the bias can be used to quantify the performance of the proposed identification algorithm in many respects. While a detailed investigation of the effect of the kernel parameters on bias (as well as on the bias/variance tradeoff) is left for future work (see the following for some numerical results in the case of a simple example), in this part it is considered the role of the sampling interval T on the quality of the computed estimates.

Indeed, in view of the expressions in (6.11)-(6.12) for the parameters of the discrete-time filters $F_{ij}(z)$, the scalar covariances in (6.18) have the following dependence on T:

$$R_{e_{p,j} e_{q,l}} = \frac{\bar{E}_{dp,j} \bar{E}_{dq,l}}{1 - G_{dj} G_{dl}} R_e \simeq \frac{T^2}{(1-T)^2} \frac{1}{1 - 1/(1-T)^2} = \frac{T}{T-2}, \qquad (6.19)$$

which clearly becomes smaller and smaller for decreasing T. Also, in view of the structure of the bias expression in (6.17), one can conclude that $\mathbb{E}[\Delta\boldsymbol{\theta}]$ decreases for decreasing T. Note that this conclusion is in agreement with the simulation results presented in Section 4, from which this effect is apparent.

6.2 Numerical example

Finally, as an example of application of the above analysis, the covariances $\mathbf{R}_{\mathbf{z}_e}$ and $\mathbf{R}_{\mathbf{z}_e e_{n+p}}$, which appear in the numerator of the bias expression in (6.17) are computed, for the case of $n = p = 1$ and $N = 2$. The kernel (3.14) is used, considering increasing values of ω ranging from 0.1 to 10 and three choices for the sampling period T, namely 10^{-3},10^{-4} and 10^{-5}. The results are summarised in Figure 6.1, in which the dependence of the two covariances (normalised to the noise variance R_e) on ω and T is depicted. As can be clearly seen from the figure, the analysis confirms the numerical results in Chapter 4 as far as the effect of T is concerned: both $\mathbf{R}_{\mathbf{z}_e}$ and $\mathbf{R}_{\mathbf{z}_e e_{n+p}}$ become negligible with respect to R_e for decreasing T; in particular, for very small values of the sampling interval, the result becomes almost insensitive to the value of ω. Furthermore, as far as the dependence on ω is concerned, the results in the figure indicate that smaller values of ω appear to be more suitable from the point of view if bias minimisation.

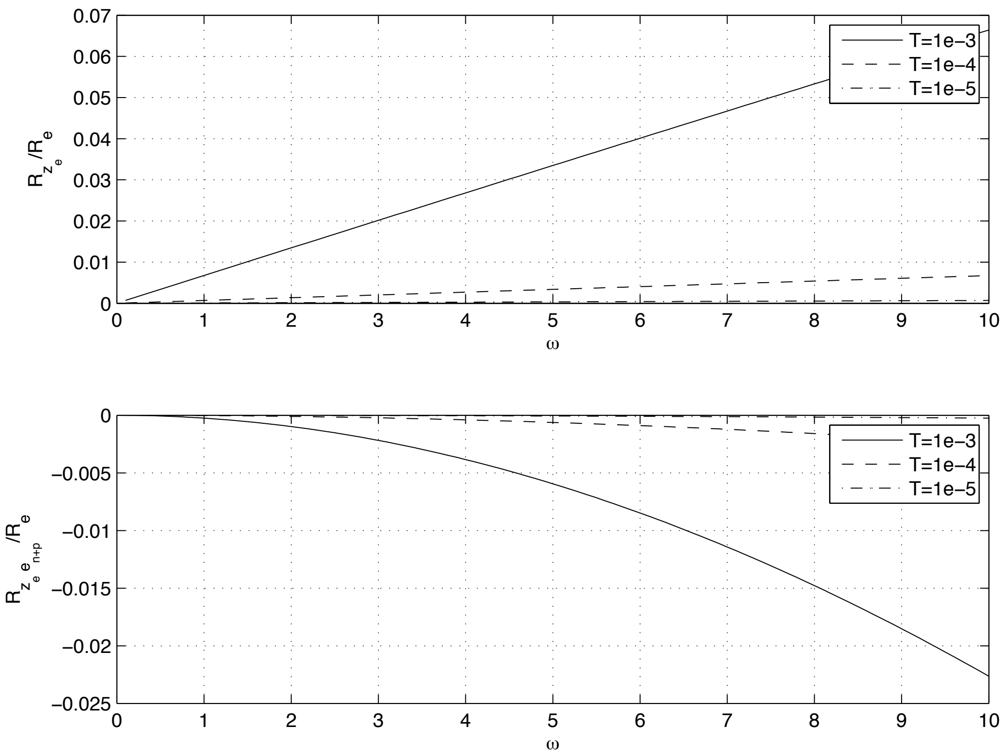


Fig. 6.1: $\mathbf{R}_{\mathbf{z}_e}$ and $\mathbf{R}_{\mathbf{z}_e e_{n+p}}$ (normalised with respect to R_e) as functions of the kernel parameter ω and of the sampling period T.

6.3 Concuding remarks

This Chapter has provided some important results in terms of the bias expression, which allows to quantify the performance of the proposed identification algorithm in many respects. It has been shown the dependence between the kernel parameter and the sampling interval and this result is very useful, since one of the main goals of this chapter was to give tuning rules for the kernel functions, therefore it is possible to choose in a suitable way the kernel parameter ω.
In the following Chapter it will be shown simulation results for the kernel-based state estimator.

Chapter 7

Simulation results

This section includes some numerical results that highlight and point out the advantages and the strength of the state estimation method presented in Chapter 5.

Consider the following third order system:

$$\begin{cases} x^{(3)}(t) = a_2 x^{(2)}(t) + a_1 x^{(1)}(t) + a_0 x(t) + b_1 u^{(1)}(t) + b_0 u(t), \\ y(t) = x(t) + \eta_y(t), \\ u(t) = v(t), \qquad\qquad t \in \mathbb{R}_{\geq 0} \end{cases} \tag{7.1}$$

System (7.1) can be written in state-space realization, using the observer-canonical form (5.2), whereas $z(0) = [z_0(0)\ z_1(0)\ z_2(0)]$, while $a_2 = -0.21$, $a_1 = -9.012$, $a_0 = -0.0901$ and $b_0 = b_1 = 1$; the output of the system, $y(\cdot)$, is affected by an unstructured perturbation term $\eta_y(\cdot)$ (addressed to as *output measurement noise*) and $u(\cdot)$ is the measured forcing input signal. The following transfer function describes the input-output behavior of (7.1):

$$G(s) = \frac{d_1 s + d_2}{s^3 + c_1 s^2 + c_2 s + c_3} \tag{7.2}$$

where $c_1 = -a_2 = 0.21$, $c_2 = -a_1 = 9.012$, $c_3 = -a_0 = 0.0901$,

$d_1 = d_2 = b_1 = b_0 = 1$. Now, to estimate the states in the BF-NK contest, we set the kernel parameters to $\omega_1 = 1$, $\omega_2 = 2$, $\omega_3 = 3$, $\varpi = 2.5$ and $\eta = n = 3$. The procedure for constructing the auxiliary signals generation system by BF-NK kernels consists in taking the derivatives of $F^{(i)}_{\omega,\varpi,\mu}(t,\tau)$, $i \in \{1,2,3\}$ of the BF-NK ((5.22)), to form the $\mathbf{E_i(t)}$ and $\mathbf{F_i(t)}$ matrices needed for the implementation of the auxiliary signal generation system $\mathcal{G}_{u,y\to\mathbf{z}_e}$. Neglecting the intermediate algebraic manipulations, we have:

$$\mathbf{E}_{\omega_1} = \begin{bmatrix} -1e^{-7.5t} + 3e^{-5t} - 3e^{-2.5t} + 1 \\ 6.5e^{-7.5t} - 12e^{-5t} + 4.5e^{-2.5t} + 1 \\ -42.5e^{-7.5t} + 48e^{-5t} - 6.75e^{-2.5t} + 1 \\ 274.625e^{-7.5t} - 192e^{-5t} + 10.125e^{-2.5t} + 1 \\ 0 \\ 0 \end{bmatrix} \tag{7.3}$$

$$\mathbf{E}_{\omega_2} = \begin{bmatrix} -1e^{-7.5t} + 3e^{-5t} - 3e^{-2.5t} + 1 \\ 5.5e^{-7.5t} - 9e^{-5t} + 1.5e^{-2.5t} + 2 \\ -30.25e^{-7.5t} + 27e^{-5t} - 0.75e^{-2.5t} + 4 \\ 166.375e^{-7.5t} - 81e^{-5t} + 0.375e^{-2.5t} + 8 \\ 0 \\ 0 \end{bmatrix} \tag{7.4}$$

$$\mathbf{E}_{\omega_3} = \begin{bmatrix} -1e^{-7.5t} + 3e^{-5t} - 3e^{-2.5t} + 1 \\ 4.5e^{-7.5t} - 6e^{-5t} - 1.5e^{-2.5t} + 3 \\ -20.25e^{-7.5t} + 12e^{-5t} - 0.75e^{-2.5t} + 9 \\ 91.125e^{-7.5t} - 24e^{-5t} - 0.375e^{-2.5t} + 27 \\ 0 \\ 0 \end{bmatrix} \tag{7.5}$$

$$\mathbf{F}_{\omega_1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1e^{-7.5t} + 3e^{-5t} - 3e^{-2.5t} + 1 \\ 6.5e^{-7.5t} - 12e^{-5t} + 4.5e^{-2.5t} + 1 \end{bmatrix} \tag{7.6}$$

$$\mathbf{F}_{\omega_2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1e^{-7.5t} + 3e^{-5t} - 3e^{-2.5t} + 1 \\ 5.5e^{-7.5t} - 9e^{-5t} + 1.5e^{-2.5t} + 2 \end{bmatrix} \tag{7.7}$$

$$\mathbf{F}_{\omega_3} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1e^{-7.5t} + 3e^{-5t} - 3e^{-2.5t} + 1 \\ 4.5e^{-7.5t} - 6e^{-5t} - 1.5e^{-2.5t} + 3 \end{bmatrix} \tag{7.8}$$

while

$$\mathbf{G}_{\omega_j} = \begin{bmatrix} -\omega_j & 0 & 0 & 0 & 0 & 0 \\ 0 & -\omega_j & 0 & 0 & 0 & 0 \\ 0 & 0 & -\omega_j & 0 & 0 & 0 \\ 0 & 0 & 0 & -\omega_j & 0 & 0 \\ 0 & 0 & 0 & 0 & -\omega_j & 0 \\ 0 & 0 & 0 & 0 & 0 & -\omega_j \end{bmatrix} \tag{7.9}$$

with $j \in \{1, 2, 3\}$.

Finally the $\mathbf{\Gamma(t)}$ matrix (see 5.28) is reported in equation (7.10)

$$\mathbf{\Gamma(t)} = \begin{bmatrix} -42.5e^{-7.5t}+48e^{-5t}-6.75e^{-2.5t}+1 & -6.5e^{-7.5t}+12e^{-5t}-4.5e^{-2.5t}-1 & -1e^{-7.5t}+3e^{-5t}-3e^{-2.5t}+1 \\ 5.5e^{-7.5t}-9e^{-5t}+1.5e^{-2.5t}+2 & 30.25e^{-7.5t}-27e^{-5t}+0.75e^{-2.5t}-4 & -1e^{-7.5t}+3e^{-5t}-3e^{-2.5t}+1 \\ 4.5e^{-7.5t}-6e^{-5t}-1.5e^{-2.5t}+3 & 20.25e^{-7.5t}-12e^{-5t}+0.75e^{-2.5t}-9 & -1e^{-7.5t}+3e^{-5t}-3e^{-2.5t}+1 \end{bmatrix} \tag{7.10}$$

7.1 Estimation in noise-free scenario

In this example we consider a noise-free scenario, where the vector of the initial conditions for system (7.1) have been set to $z(0) = [1\ -10\ 1]$, while the forcing input has been chosen as a sum of sinusoids $u(t) = v(t) = 10\sin(10t) + \sin(2t)$, depicted in 7.1.
In Figure 7.2 it is shown that the theoretical istantaneity of the method has effectiveness also in numerical simulations, in fact the states are correctly estimate by the proposed method with remarkable precision and with negligible delay (the estimator has been initialized to zero until time $t = 0.1\ s$). It is worth noting that the proposed BF-NK estimator, beyond fast convergence, is characterized by guaranteed internal stability, hence re-initialization is not required.

We point out that no high-gain output injection has been performed by

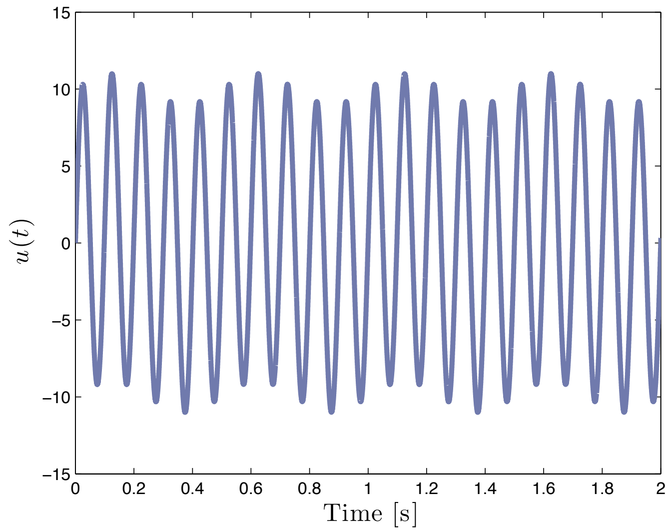
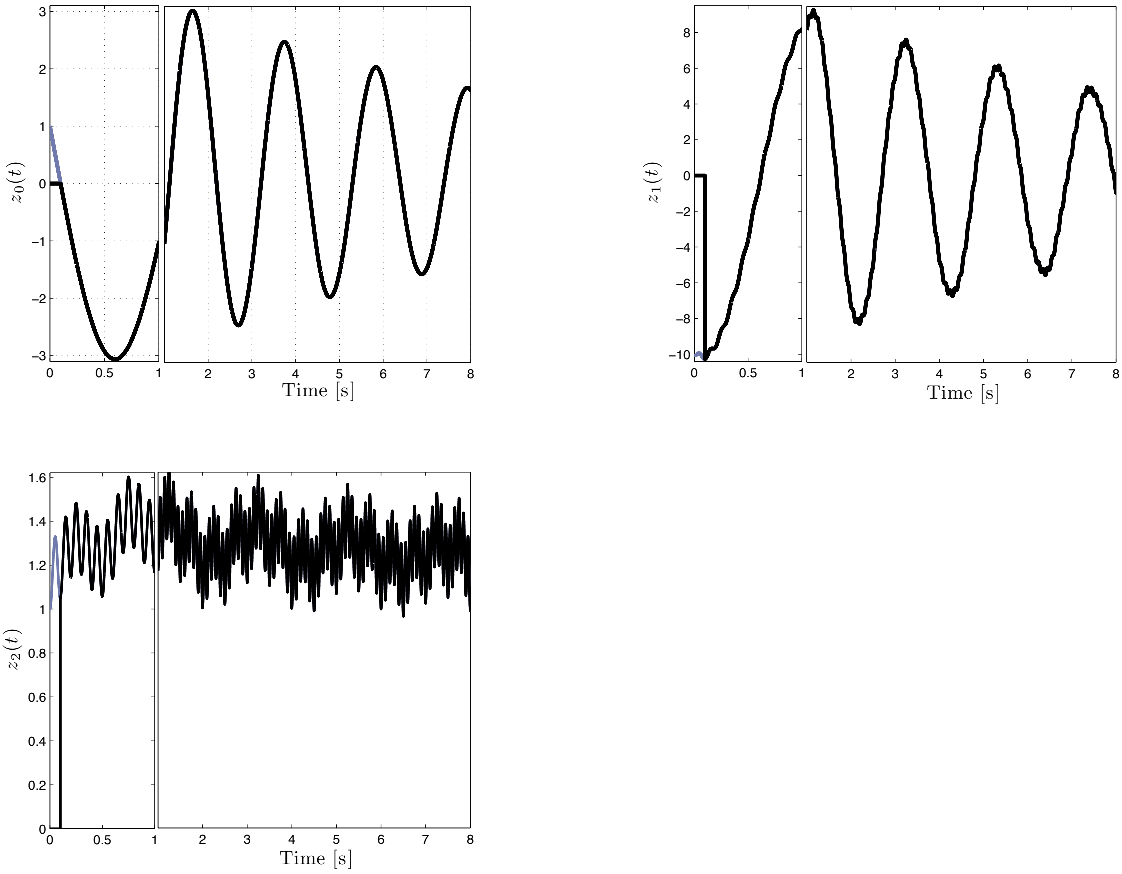


Fig. 7.1: Trends of the measured signals $u(t)$ used for the estimation in noise-free conditions.

the two methods. In this respect, a further simulation has been carried out in noisy conditions.



(c)

Fig. 7.2: The states of system (7.1) (gray) and estimated by the BF-NK estimator (black) in noise-free conditions. The estimates of the observed states are exact after an arbitrarily small finite-time, depending on the inversion time-instant of the time-varying $\mathbf{\Gamma(t)}$ matrix (equation (5.28)). To avoid singularities due to numerical precisions, we have choosen to invert matrix $\mathbf{\Gamma}$ from time $t = 0.1$ s in the example.

7.2 Estimation with Unstructured Measurement Perturbations

In this example, the additive output measurement noise $\eta_y(\cdot)$ has been simulated as a uniformly distributed random signal taking values in the interval $[-0.2, 0.2]$. The perturbed signal used for parameter estimation is depicted in Fig. 7.3.

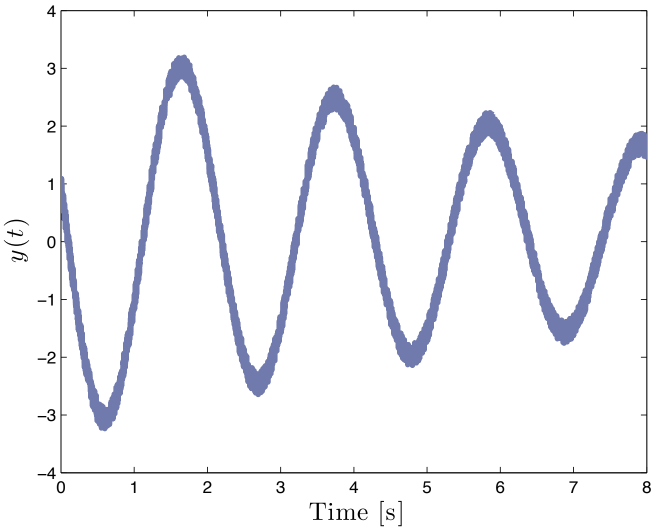
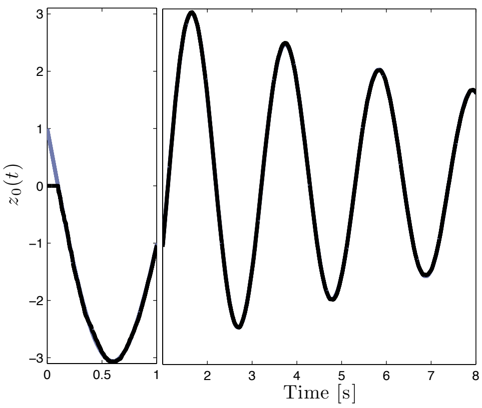
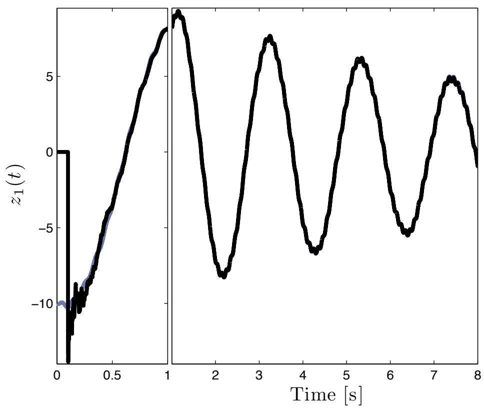


Fig. 7.3: Trends of the measured signal $y(t)$ used for the estimation in noise conditions.

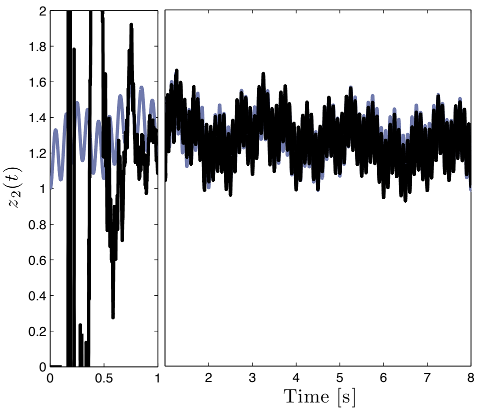
As can be seen in Fig. 7.4, the BF-NK estimator shows good robustness against the output noise and the estimated parameters converge to a neighborhood of the true values. The characterization, in both deterministic and stochastic settings, of the behavior of the BF-NK estimator in the presence of measurement noise (Chapter 6) allowed to determine tuning rules for the ω parameters such that the noise effect is minimized.



(a)



(b)



(c)

Fig. 7.4: The states of system 7.1 (gray) and estimated by the BF-NK estimator (black) in noise conditions converge to the true values.

Chapter 8

Conclusions

In this work, a comprehensive kernel-based system identification architecture, suitable for identification of continuous-time linear systems, has been designed. The motivations for this thesis work are the renewed interest given to continuous-time systems, thanks to their easier mathematical implementation and its practical advantages compared with the discrete-time systems. Besides, the incresed complexity in modern systems implies the need for novel tools, able to work in different contexts where the lack of data is often the real scenario; therefore the unavailability of the time-derivatives of dynamical system and the unknown of initial conditions is a typical example of these cases; to address these situations, it is necessary to have robust and adaptable tools. That's why we design a comprehensive identification technique, taking into account several aspects:

- design a novel class of kernels allowing to get rid of the influence of the unknown initial conditions; this aspect is one of the main results arising from this work;
- provide a technique that combines fast convergency properties with a guarantee internal stability; this can be very useful, because it encloses the main advantages of two different techniques available in literature: indeed, SVF methods show nice features in their asymptotic behavior, while the transient is usually overlooked, and IM methods have good behaviors in the transient (*i.e.*, whence there is a reduced dataset), while asymptotically, due to windup issues and moving horizon window, is of little use.
- obtain a robust methodology with respect to stochastic noises to provide the consistency of the estimation method.

All of these aspects have been addressed along this work, exploiting well-knwon mathematical tools. It has been shown how, starting from Volterra linear integral operators, it has been possible design kernel functions such that the estimation system was solvable, linear and with guarantee internal stability properties. Besides, it has been shown that, designing these kernel functions in such a way that several properties are satisfied, it is possible to affect on the convergence time of the estimation and obtain a non-asymptotic estimation. Finally, it has been provided tuning rules for the kernel parameters based on the bias analysis of the estimator providing theoretical arguments that confirm the effectiveness of the estimation method presented.

8.1 Future developments

As a future work, it will be investigate the possibility to extend this method to some classes of nonlinear dynamical systems. Nonlinear continuous-time systems can be represented by a state-space description, or alternatively by an input-output description.

A state-space description of a continuous-time nonlinear system has the form:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= g(x(t), u(t)) \end{aligned}$$

where $x(t)$, $u(t)$ and $y(t)$ are respectively the state, the input and the output of the system and f and g are nonlinear vector functions.

State-space systems are more attractive for dealing with multivariable inputs and outputs; as argued by Rivals and Personnaz [101], state-space systems are likely to require fewer parameters, especially for multivariable systems. Because of these reason, state-space models are often preferred to input-output models.

The particular choice of the functions f and g for the state-space model, determines the structure of the nonlinear model. Of course many possibilities exist. In general, two approaches can be distinguished. The first approach is to choose a simple structure for the model, such that it has certain nice properties, is easy to analyze, and leads to computationally attractive identification methods. The main drawback is that simple structures often represent a limited class of nonlinear systems. Therefore, the second approach aims at choosing the model structure such that the model can approximate a large class of nonlinear systems. The disadvantage is that this often leads to complicated identification. Examples of the first approach are the Hammerstein model [102], while the second approach include hinging hyperplanes models [103].

Moreover, ongoing research aims at extend the methodology to Multiple-Input-Multiple-Output (MIMO) systems. In general terms, this topic can be viewed as the problem of finding mapping between the available input-output data sequences and unknown parameters in a user defined class of models.

Another open problem in kernel-based system identification is the derivation of norm bounds on the appoximation errors of the methods that were discussed: a solution to this kind of problem would allow a direct comparision of the approximation errors. However deriving this bound is not trivial, becuase it involves products of matrices with varying dimensions. Another problem is the derivation of conditions on the input system that ensure that we are closely related to the persistency of excitation. Since the data matrices contain products of time-lagged versions of the input, output and parameters, it has necessary to provide that this higher-order moment matrices have full row rank for its invertibility w.r.t. the input function design.

Appendix

Composition of Volterra operators

We will show that the composition of two integral Volterra operators yields, in turn, a Volterra integral operator. This result descends from the composition property of Fredholm operators (see [83]). The derivation is non-trivial due to the necessity of considering explicitely the integration over finite domains.
Suppose that V_{K_h} and V_{K_g} are two Volterra operators induced by the $\mathcal{HS}$ kernels K_h and K_g respectively. The composition of the two operators results in a double integral:

$$\left[V_{K_h}\left[V_{K_g}\right]\right](t) = \int_0^t K_h(t,\sigma)\left(\int_0^\sigma K_g(\sigma,\tau)x(\tau)d\tau\right)d\sigma.$$

By introducing the Fredholm extension of the Volterra kernel: $\tilde{K}_g(\sigma,\tau) \triangleq K_g(\sigma,\tau)H(\tau)H(\sigma-\tau)$, it is possible to extend the limit of the inner integral from σ to t:

$$\begin{aligned}
\left[V_{K_h}\left[V_{K_g}\right]\right](t) &= \int_0^t K_h(t,\sigma)\left(\int_0^t \tilde{K}_g(\sigma,\tau)x(\tau)d\tau\right)d\sigma \\
&= \int_0^t\int_0^t K_h(t,\sigma)\tilde{K}_g(\sigma,\tau)x(\tau)d\tau d\sigma \\
&= \int_0^t\left(\int_0^t K_h(t,\sigma)\tilde{K}_g(\sigma,\tau)d\sigma\right)x(\tau)d\tau \\
&= \int_0^t\left(\int_\tau^t K_h(t,\sigma)K_g(\sigma,\tau)d\sigma\right)x(\tau)d\tau \\
&= \int_0^t (K_h \bullet K_g)(t,\tau)x(\tau)d\tau
\end{aligned} \tag{8.1}$$

where the kernel of the composed integral operator can be thus obtained by the *kernel composition integral* $(\cdot \bullet \cdot)$, defined as:

$$(K_h \bullet K_g)(t,\tau) \triangleq \int_\tau^t K_h(t,\sigma)K_g(\sigma,\tau)d\sigma. \tag{8.2}$$

The causality of the Volterra operator has the following important implication:

$$(K_h \bullet K_g)(t,t) = 0, \;\; \forall t \in \mathbb{R}_{\geq 0}.$$

Moreover, if for some $i \in \mathbb{N}$, $K_g^{(i)}(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}$, then it is trivial to prove that

$$(K_h \bullet K_g)^{(i)}(t,0) = 0, \;\; \forall t \in \mathbb{R}_{\geq 0}.$$

Changing the order of summation in nested indexed sums

In the following, we will report some technical results used in various parts of the manuscript. We will exploit the Iverson's Bracket notation (see [104] and the reference therein) and the bracket multiplication property to change the order of summation in nested indexed sums. Let $g_{i,j} \in \mathbb{R}$ denote the elements of a double indexed array, with $i, j \in \mathbb{Z}$, and let $m, n \in \mathbb{Z}_{\geq 1}$ be finite integers, with $m \leq n$. Then it holds that:

$$\begin{aligned}
\sum_{i=0}^{n-1}\sum_{j=0}^{i-1} g_{i,j} &= \sum_{i}\sum_{j}[0 \leq i \leq n-1][0 \leq j \leq i-1]g_{i,j} \\
&= \sum_{i,j}[0 \leq i \leq n-1][1 \leq j+1 \leq i]g_{i,j} \\
&= \sum_{i,j}[1 \leq j+1 \leq i \leq n-1]g_{i,j} \\
&= \sum_{j,i}[1 \leq j+1 \leq n-1][j+1 \leq i \leq n-1]g_{i,j} \\
&= \sum_{j}\sum_{i}[0 \leq j \leq n-2][j+1 \leq i \leq n-1]g_{i,j} \\
&= \sum_{j=0}^{n-2}\sum_{i=j+1}^{n-1} g_{i,j}
\end{aligned} \tag{8.3}$$

$$
\begin{aligned}
\sum_{j=0}^{n-2}\sum_{i=j+2}^{n} g_{i,j} &= \sum_{j}\sum_{i}[0 \le j \le n-2][j+2 \le i \le n]g_{i,j} \\
&= \sum_{j,i}[2 \le j+2 \le n][j+2 \le i \le n]g_{i,j} \\
&= \sum_{i,j}[2 \le j+2 \le i \le n]g_{i,j} \\
&= \sum_{i,j}[2 \le i \le n][2 \le j+2 \le i]g_{i,j} \\
&= \sum_{i}\sum_{j}[2 \le i \le n][0 \le j \le i-2]g_{i,j} \\
&= \sum_{i=2}^{n}\sum_{j=0}^{i-2} g_{i,j}
\end{aligned}
\tag{8.4}
$$

$$
\begin{aligned}
\sum_{j=0}^{m-2}\sum_{i=j+2+n-m}^{n} g_{i,j} &= \sum_{j}\sum_{i}[0 \le j \le m-2][j+2+n-m \le i \le n]g_{i,j} \\
&= \sum_{j,i}[2+n-m \le j+2+n-m \le n][j+2+n-m \le i \le n]g_{i,j} \\
&= \sum_{i,j}[2+n-m \le j+2+n-m \le i \le n]g_{i,j} \\
&= \sum_{i,j}[2+n-m \le i \le n][2+n-m \le j+2+n-m \le i]g_{i,j} \\
&= \sum_{i}\sum_{j}[2+n-m \le i \le n][0 \le j \le i-2-n+m]g_{i,j} \\
&= \sum_{i=2+n-m}^{n}\sum_{j=0}^{i-2-n+m} g_{i,j}
\end{aligned}
\tag{8.5}
$$

Bibliography

[1] L.A.Zadeh, "From circuit theory to system theory," in *Proc. IRE 50*, 1962, pp. 856 – 865.

[2] S. Haykin, *Neural networks: a comprehensive foundation.* Upper Saddle River, New Jersey: Prentice Hal, 2004.

[3] G. Pin, A.Assalone, M. Lovera, and T. Parisini, "Kernel-based non-asymptotic parameter estimation of continuous-time systems," in *Proc. 51st IEEE Conference on Decision and Control*, Maui, USA, 2012.

[4] G. Pin, M. Lovera, A. Assalone, and T. Parisini, "Kernel-based non-asymptotic state estimation for linear continuous-time systems," in *Proc. 2013 American Control Conference*, 2013.

[5] G.Pin, A.Assalone, M.Lovera, and T.Parisini, "Kernel-based non-asymptotic parameter estimation of continuous-time linear systems," *Submitted IEEE Trans. on Automatic Control*, 2013.

[6] R. Brockett, *Finite Dimensional Linear Systems.* New York: Wiley, 1970.

[7] K.Aström and P.Eykhoff, "System identification - a survey," *Automatica*, vol. 7, 1971.

[8] K. Aström and B. Wittenmark, *Computer controlled systems: theory and design.* Upper Saddle River, New Jersey: Prentice Hal, 1984.

[9] T.Bohlin, "On the maximum likelihood method of identification," *IBM Journal of research and development*, vol. 14, pp. 41–51, 1970.

[10] T. Bohlin, "On the problem of ambiguities in maximum likelihood identification," *Automatica*, vol. 7, no. 2, pp. 199–210, 1971.

[11] B. Ho and R. Kalman, "Effective construction of linear state-variable models from input/output data," in *Proc. of the Annual Allerton Conf. on Cricuit and System Theory*, Monticello, Illinois, 1965.

[12] G. Box and G. Jenkins, *Time series analysis, forecasting and control.* Oackland: Holden-Day, 1976.

[13] P. Eykhoff, *System identification.* Wiley, 1974.

[14] L. Ljung, *System Identification - Theory for the user.* Englewood Cliffs: Prentice-Hall, 1987.

[15] J. Norton, *An introuction to identification.* London: Academic Press, 1986.

[16] T. Söderström and P. Stoica, *System Identification.* New York: Prentice-Hall, 1989.

[17] C. Lawson and R.J.Hanson, *Solving least squares problems.* Philadelphia, Pennsylvania: SIAM, 1995.

[18] M.Verhaegen and P. Dewilde, "Subspace model identification part 1: the output-error-state-space model identification claass of algorithmspower system frequency estimation using least mean square technique," *International Journal of Control*, vol. 56, no. 5, pp. 1187–1210, 1992.

[19] P. V. Overschee and B. D. Moor, *Subspace identification for linear systems: theory, implementation, applicatios.* Dordrecht, NL: Kluver Academic Publisher, 1996.

[20] T. Katayama, *Subspace methods for system identification: a realization approach.* London: Springer-Verlag, 2005.

[21] F. Eng and F. Gustafsson, "Identification with stochastic sampling time jitter," *Automatica*, vol. 44, no. 2, pp. 637–646, 2006.

[22] F.Marvasti, *Nonuniform sampling: theory and practice.* Dordrecht, NL: Kluver Academic Publisher, 2001.

[23] R.Johansson, "Continuous-time model identification and state estimation using non-uniformly sampled data," in *Proc. of the 19th IEEE Symposium on Mathematical Theory of Networks and Systems*, Budapest, 2010, pp. 347–354.

[24] K. Aström, "On the choice of sampling rates in parametric identification of time series," *Information Sci.*, vol. 1, no. 1, pp. 273–278, 1969.

[25] P. Young, "Parameter estimation for continuous-time models- a survey," *Automatica*, vol. 17(1), pp. 23–39, 1981.

[26] H.Unbehauen and G.P.Rao, "Continuous-time approaches to system identification- a survey," *Automatica*, vol. 26(1), pp. 23–35, 1990.

[27] N.K.Sinha and G.P.Rao, *Identification of Continuous-Time Systems. Methodology and Computer Implementation*, Kluwer, Dordrecht, NL, 1991.

[28] H.Garnier, M.Mensler, and A.Richard, "Continuous-time model identification from sampled data: implementation issues and performance evaluation," *Int. J. Control*, vol. 76, no. 13, pp. 1337–1357, 2003.

[29] H.Garnier and M.Mensler, "The CONTSID toolbox: A matlab toolbox for continuous-time system identification," in *Proc. 12th IFAC Symp. System Identification*, 2000.

[30] H.Garnier, M.Gilson, and E.Huesestein, "Developments for the matlab CONTSID toolbox," in *Proc. IFAC Symp. System Identification-American Control Conference*, 2003.

[31] T.Sugie and T.Ono, "An iterative learning control law for dynamical systems," *Automatica*, vol. 27, no. 4, pp. 729–732, 1991.

[32] F.Sakai and T.Sugie, "Continuous-time system identification based on iterative learing control," in *Proc. 16th IFAC World Congress*, 2005.

[33] M.Campi, T.Sugie, and F.Sakai, "An iterative identification method for linear continuous-time systems," *IEEE Trans. on Automatic Control*, vol. 53, no. 7, pp. 1661–1669, 2008.

[34] J. Sjöberg, Q.Zhang, L.Ljung, A.Benveniste, B. Delyon, P. Glorrenec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.

[35] S. Billings, "Identification of nonlinear systems: a survey," *IEE Proceedings Part D:Control Theory and Applications*, vol. 127, no. 6, pp. 272–285, 1980.

[36] S.Chen and S. Billings, "Representations of non-linear systems: the narmax model," *International Journal of Control*, vol. 49, no. 3, pp. 1013–1032, 1989.

[37] K. Narenda and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Networks*, vol. 1, no. 4, pp. 4–27, 1990.

[38] R. Haber and H. Unbehauen, "Structure identification of nonlinear dynamic system. a survey on input/output approaches," *Automatica*, vol. 26, no. 4, pp. 651–677, 1990.

[39] S. Billings, H. Jamaluddin, and S.Chen, "Properties of neural networks with applications to modelling non-linear dynamical systems," *International Journal of Control*, vol. 55, no. 1, pp. 193–224, 1992.

[40] A. Juditsky, H. Hjalmarsson, A.Benveniste, B. Delyon, J. Sjöberg, L.Ljung, and Q.Zhang, "Nonlinear black-box modeling in system identification: mathematical foundations," *Automatica*, vol. 31, no. 12, pp. 1725–1750, 1995.

[41] G. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, pp. 533–580, 2001.

[42] I. Landau, *System Identification and Control Design*. New York: Prentice-Hall, 1990.

[43] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Prentice-Hall, 1994.

[44] H. Garnier and L. Wang, Eds., *Identification of Continuous-time Models from Sampled Data*. Advances in Industrial Control: Springer, 2008.

[45] P. Young, *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Pratictioner*. Berlin: Springer-Verlag, 2011.

[46] L. Belkoura, J. Richard, and M. Fliess, "Parameters estimation of system with delayed and structured entries," *Automatica*, vol. 39, no. 2, pp. 291–298, 2009.

[47] M. Mboup, C. Join, and M. Fliess, "Numerical differentiation with annihilators in noisy environment," *Numerical Algorithms*, vol. 50, no. 4, pp. 439–467, 2009.

[48] M. Fliess, C. Join, and H. Sira-Ramírez, "Robust residual generation for linear fault diagnosis: An algebraic setting with examples," *Int. J. Control*, vol. 77, no. 4, pp. 1223–1242, 2004.

[49] M.Fliess, C.Join, and H.Sira-Ramírez, "Non-linear estimation is easy," *Int. J. Model., Identification and Control*, vol. 4, no. 1, pp. 12–27, 2007.

[50] M. Fliess and H. Sira-Ramírez, "An algebraic framework for linear identification," *ESAIM Control Optim. Calc. Variat.*, vol. 9, pp. 151–168, 2003.

[51] H.Sira-Ramírez and S.Agrawal, *Differentially flat systems*. New York: Marcel Dekker, 2004.

[52] H. Sira-Ramírez and M. Fliess, "An algebraic state estimation approach for the recovery of chaotically encrypted messages," *Int. J. Bifurcation and Chaos*, vol. 16, no. 2, pp. 295–309, 2006.

[53] J. Linares-Flores, J. Reger, and H. Sira-Ramírez, "Load torque estimation and passivity-based control of a boost-converter/DC-motor combination," *IEEE Trans. Control Systems Technology*, vol. 18, no. 6, pp. 1398–1405, 2010.

[54] J. Trapero, H. Sira-Ramírez, and V. Batlle, "An algebraic frequency estimator for a biased and noisy sinusoidal signal," *Signal Process.*, vol. 87, no. 6, pp. 1188–1201, 2007.

[55] L. Belkoura, C. Join, and M. Mboup, "Algebraic change-point detection," *Appl. Algebra Eng. Commun. Comput.*, vol. 21, no. 2, pp. 131–143, 2010.

[56] P. C. Young, *Process parameter estimation and self adaptive control.* New York: Plenum Press, 1965.

[57] D. Q. Mayne, "A method for estimating discrete-time transfer functions," in *Advances in Computer Control, Second UKAC Control Convention*, University of Bristol, 1967.

[58] K.Wong and E. Polak, "Identification of linear discrete-time systems using the instrumental variable approach," *IEEE Transaction on Automatic Control*, vol. AC-12, pp. 707–718, 1967.

[59] P. C. Young, "Regression analysis and process parameter estimation: a cautionary message," *Simulation*, vol. 10, pp. 125–128, 1968.

[60] P. Young, "The use of linear regression and related procedures for the identification of dynamic processes," in *Proc. of the 7th 2001 IEEE Symposium on Adaptive Processes*, New York, 1969.

[61] P. C. Young, "Some observations on instrumental variable methods of time-series analysis," *International Journal of Control*, vol. 23, pp. 593–612, 1976.

[62] P. Young and A. Jakeman, "Refined instrumental variable methods of time-series analysis: part i, siso systems," *International Journal of Control*, vol. 29, pp. 1–30, 1979.

[63] A. Jakeman and P. Young, "Refined instrumental variable methods of time-series analysis: part ii, multivariable systems," *International Journal of Control*, vol. 29, pp. 621–644, 1979.

[64] P. Young and A. Jakeman, "Refined instrumental variable methods of time-series analysis: part iii, extensions," *International Journal of Control*, vol. 31, pp. 741–764, 1980.

[65] P. C. Young, *Recursive estimation and time-series analysis.* Berlin: Springer-Verlag, 1984.

[66] P. Young, H. Garnier, and M. Gilson, "An optimal instrumental variable approach for identifying hybrid continuous-time box-jenkins models," in *Proc. of the 14th Symposium on System Identification*, Newcastle, Australia, 2006, pp. 225–230.

[67] C. Chen, *Linear system theory and design.* Oxford University Press, 1998.

[68] J. Gertler, "Survey of model-based failure detection and isolation in complex plant," *IEEE Control System Magazine*, December 1998.

[69] G.Ellis, *Observers in Control Systems: a Practical Guide.* Academic Press, 2002.

[70] T.Umeno and Y.Hori, "Robust speed control of dc servomotors using modern two degrees-of-freedom controller design," *IEEE Transactions on Industrial Electronics*, vol. 38, pp. 1309–1314, 1991.

[71] Y.Hori and K.Shimura, "Position/force control of multi-axis robot manipulator based on the tdof robust servo controller for each joint," in *Proc. American Control Conference*, 1992.

[72] D. Luenberger, "Observers for multivariable systems," *IEEE Transactions on Automatic Control*, vol. 11, pp. 190–196, 1966.

[73] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *J. of Basic Engineering., Trans. ASME, Series D*, vol. 83, pp. 95–108, 1961.

[74] S. Han, W. Kwon, and P. Kim, "Receding-horizon unbiased FIR filters for continuous-time state-space models wihout a priori state informations," *IEEE Transactions on Automatic Control*, vol. 46, pp. 766–770, 2001.

[75] W. Byrski, "The survey of exact and optimal state observers in hilbert spaces," in *Proc. of European Control Conference*, Cambridge, UK, 2003.

[76] R. Engel and G. Kresselmeier, "A continuous-time observer which converges in finite-time," *IEEE Transactions on Automatic Control*, vol. 47, pp. 1202–1204, 2002.

[77] A. Medvedev and T. Toivonen, “Feedforward time-delay structures in state estimation: finite memory smoothing and continuous deadbeat observers,” *IEE Proceedings of Control Theory and Applications*, vol. 141, pp. 121–129, 1994.

[78] I. Hasakara, U.Özgüner, and V. Utkin, “On sliding mode observers via equivalent control approach,” *Int. J. Contr.*, vol. 71, no. 6, pp. 1051–1067, 1998.

[79] T. Raff and F. Allgöwer, “Observers with impulsive dynamical behavior for linear and nonlinear continuous-time systems,” in *Proceedings of the IEEE Conference on Decision and Control*, New Orleans, 2007, pp. 4287–4292.

[80] T.Raff and F. Allgöwer, “An observer that converges in finite time due to measurement-based state updates,” in *Proc. of the 17th IFAC World Congress*, Seoul, Korea, 2008, pp. 2693–2695.

[81] C.Curduneanu, *Integral Equations and Applications*. Cambridge University Press, 1991.

[82] I. Bronsthein, K. Semendyayev, G.Musiol, and H.Muehlig, *Handbook of Mathematics*. Springer, 2005.

[83] T. Burton, *Volterra Integral and Differential Equations*. Elsevier, 2005.

[84] A. Feintuch and R. Saeks, *System Theory. A Hilbert Space Approach*. New York: Academic Press, 1982.

[85] R.Kress, *Linear integral equations*. Berlin: Springer, 1999.

[86] A. Kolmogorov and S.Fomin, *Elements of the theory of functions and functional analysis*. New York: Courier Dover Publications, 1999.

[87] M. Bergamasco and M. Lovera, “Continuous-time predictor-based subspace identification using Laguerre filters,” *IET Control Theory and Applications*, vol. 5, no. 7, pp. 856–867, 2011.

[88] Y. Ohta, “Stochastic system transformation using generalized orthonormal basis functions with applications to continuous-time system identification,” *Automatica*, vol. 47, no. 5, pp. 1001–1006, 2011.

[89] C. García-Rodríguez, J. Cortés-Romero, and H.Sira-Ramírez, “Algebraic identification and discontinuous control for trajectory traking in a perturbed 1-DOF suspension system,” *IEEE Trans. Industrial Electronics*, vol. 56, no. 9, pp. 3665–3674, 2009.

[90] G. P. Rao and H. Garnier, "Numerical illustrations of the relevance of direct continuous-time model identification," in *15th Triennal IFAC World Congress on Automatic Control*, Barcelona, Spain, 2002.

[91] V. Laurain, H. Garnier, M. Gilson, and P. Young, "Refined instrumental variable method for identification of hammerstein continuous-time box-jenkins models," in *Proc. of IEEE Conference on Decision and Control*, Cancun, 2008, pp. 1386–1391.

[92] L. Ljung, "Initialisation aspects for subspace and output-error identification methods," in *Proc. European Control Conference, ECC 03*, J. Maciejowski, Ed., Cambridge, UK, Sept. 2003.

[93] H.Unbehauen and G.P.Rao, "A review of identification in continuous-time systems," *Annual Reviews in Control*, vol. 22, pp. 145–171, 1990.

[94] P.Young, "The refined instrumental variable method: unified estimation of discrete and continuous-time transfer function models," *Journal européen des systemés automatisés*, vol. 42, no. 2, 2008.

[95] M. Shinbrot, "On the analysis of linear and non linear systems," *Transactions on the ASME*, vol. 79, pp. 547–552, 1957.

[96] H.Garnier, M.Gilson, and V.Laurain, "The CONTSID toolbox for matlab: extensions and latest developments," in *Proc. 15th IFAC Symp. on System Identification*, 2009.

[97] R. Bracewell, *The Hartley Transform.* New York: Oxford University Press, 1986.

[98] P. C. Young, H.Garnier, and M. Gilson, *An optimal instrumental variable approach for identifying hybrid continuous-time Box-Jenkins models.* Berlin: Springer-Verlag, 2008.

[99] D.Luenberger, "Observers for multivariable systems," *IEEE Trans. on Automatic Control*, vol. 11, no. 2, pp. 190–197, 1966.

[100] L.Ljung and A.Willis, "Issues in sampling and estimating continuous-time models with stochastic disturbances," *Automatica*, vol. 46, no. 5, pp. 925–931, 2010.

[101] I. Rivals and L. Personnaz, "Black-box modeling with state-space neural networks," in *Neural Adaptive Control Technology I*, 1996, pp. 237–264.

[102] I. Leontaritis and S. Billings, "Input-output parametric models for non-linear systems part II: stochastic non-linear systems," *International Journal of Control*, vol. 41, no. 2, pp. 329–344, 1985.

[103] P.Pucar and J. Sjöberg, "On the hinge-finding algorithm for hinging hyperplanes," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1310–1319, 1998.

[104] R.Graham, D.Knuth, and O.Patashnik, *Concrete Mathematics.* Addison-Wesley Publishing Company, 1994.