Kernel-Based Generalized Least Squares for Joint State and Parameter Estimation in LTI Systems

Santosh Devapati

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

August 2020

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

2020/08/21

Abstract

A Stochastic multiple regression approach for parameter and state estimation in Reproducing Kernel Hilbert Space (RKHS) is presented in this thesis. It begins with the understanding and derivation of double sided kernel representation for a fourth order linear system and proceeds into discussing and developing methods for state and parameter estimation from single noisy realizations of the system output on a time interval $[a, b]$.The multiple linear regression model does not satisfy the assumptions of the Gauss-Markov theorem in that the random regressor has a regression error, which is heteroskedastic. These complications do not impede achieving high accuracy of estimation. A recursive version of a feasible generalized least squares with covariance weighting is employed to attenuate adverse effects due to heteroskedasticity. Once the parameters are estimated the output is reconstructed by projection onto the span of fundamental solutions and this in turn is used to reconstruct the time derivatives of the system output.

Résumé

Une approche de régression multiple stochastique pour l'estimation des paramètres et des états dans l'espace de reproduction du noyau de Hilbert (RKHS) est présentée dans cette thèse. Il commence par la compréhension et la dérivation de la représentation du noyau double face pour un système linéaire du quatrième ordre et passe à la discussion et au développement de méthodes d'estimation d'état et de paramètres à partir de réalisations bruyantes uniques de la sortie du système sur un intervalle de temps [a, b]. Le modèle de régression linéaire multiple ne satisfait pas aux hypothèses du théorème de Gauss-Markov en ce que le régresseur aléatoire a une erreur de régression, qui est hétéroscédastique. Ces complications n'empêchent pas d'atteindre une grande précision d'estimation. Une version récursive des moindres carrés généralisés réalisables avec pondération de la covariance est utilisée pour atténuer les effets négatifs dus à l'hétéroscédasticité. Une fois les paramètres estimés, la sortie est reconstruite par projection sur la durée des solutions fondamentales et celle-ci est à son tour utilisée pour reconstruire les dérivées temporelles de la sortie du système.

Acknowledgment

I would like to express my gratitude to several individuals who constantly helped me throughout my academic life at McGill.

Firstly I thank Professor Hannah Michalska, who is my supervisor. She gave me the right direction patiently through the entire research and helped to shape this research thesis elegantly.

I would like to thank my fellow research students Surya Kumar Devarajan and Shantanil Bagchi for their feedback, cooperation, and of course friendship. I am grateful to my parents,my brother,all my family and friends who have been a great encouragement.

Last but not the least, I would like to thank Mr.A.Sudhir and Mr.G.Srinivasa Rao and all my NFCL colleagues who have encouraged me to pursue my masters. My Deepest gratitude towards Sindhura Vajrala for her constant support and love throughout writing this thesis and my life in general.

Preface

This is to declare that the work presented in this document was completed and carried out by Santosh Devapati and is part of a collaborative work of three member team guided by Professor Hannah Michalska. The theory of representation of n-th order system using kernels in RKHS was derived by Debarshi Patanjali Ghoshal, PhD scholar in the research group and the representation for fourth order system output and its derivatives was verified and coded by me and also used in the implementation of the cost function for this thesis work. The theoretical background for the RKHS approaches for optimization is based on the research notes by Professor Hannah Michalska, which is duly acknowledged.

List of Figures

List of Tables

List of Acronyms

RKHS	Reproducing Kernel Hilbert Space
KS	Kolmogorov Smirnov
LTI	Linear Time Invariant
RMSD	Root Mean Square Difference
SISO	Single Input Single Output
OLS	Ordinary Least Squares
RLS	Recursive Least Squares
DSK	Double Sided Kernel
GLS	Generalised Least Squares
FGLS	Feasible Generalised Least Squares
BLUE	Best Linear Unbiased Estimator

Chapter 1

Introduction

A control system model is a mathematical representation of the essential characteristics of an existing control system. A control system uses a controller which provides the corrective action eliminating the error to achieve a desired output. It is a known fact that a controller is of two types: open loop and closed loop controller. An *open-loop* control system is a system which cannot correct variations in the output because it does not employ feedback. Whereas a *closed loop* system is a type of system which employs a feedback loop. An error detector compares a signal, a function of the output, obtained from a sensor, with a reference input. The difference between these signals is used by the controller to determine a control action and reduce the error.The difference between the desired output and the current output generates a control law so as to change the state or the parameters of the system to bring the system output closer to the desired output. A general control system employing feedback is represented as below [26].

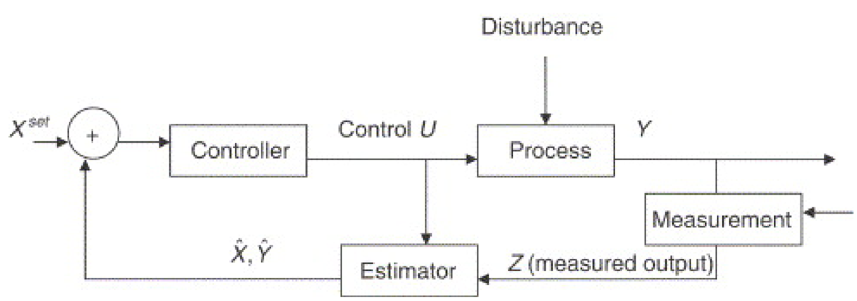

Figure 1.1 Closed Loop Controller with State Estimator

1.1 Parameter Estimation and State Estimation in Linear Systems

"When a system model can be defined by a finite number of variables and parameters, it is called a parametric model.To implement a parametric model-based controller, it is necessary to know precisely the structure of the model of the system and its associated parameters. Therefore, if parameters are initially unknown, the process of parameter identification is quite important for the design of the control system" [26].Identification of parameters have important applications in system identification, system control, and system analysis.
"The minimum set of variables whose current values along with the values of the input signals in future can predict the future behavior of the system are state variables" [26].For a System to generate a desired control action in order to reduce the error the understanding of all systems states are usually required. The system is fed with measurement noise and input disturbances hence it is expected that the controller has to be less sensitive to these external signals therefore for a large system, it is not possible to measure the state variables. An estimator needs to be designed that can estimate the states from both the output and input. The problem of estimator design for an observable system has been extensively dealt by mathematicians especially, Kalman and Luenberger [19], [20], [21], [22].
The framework of parameter and state estimation based on an algebraic method been introduced by M. Fliess and H. Sira-Ramirez [9]. It is based on differential algebra and operational calculus. The algebraic method is non-asymptotic: the solutions are obtained by exact algebraic formulas, one to obtain estimates in finite time. The desired parameters are expressed as a function of integrals of the measured outputs and inputs of the system [7]. It does not need any statistical knowledge of the noise . Moreover, the estimator is able to treat identification in the presence of bias and of structured perturbations.Noise attenuation is attempted by repeated integration and shaping the annihilator functions used in eliminating the effect of the initial conditions. However, this method requires frequent re-initialization when used forward in time, and its noise rejection properties were characterized as non-standard; see Fliess (2006). These methods are sensitive to measurement noise.
Algebraic state and parameter estimation of linear systems based on a special construction of a forward-backward kernel representation of linear differential invariants are extended to handle large noise in output measurement [13] .With the knowledge of the characteristic equation of the system the state equations are replaced with an output reproducing property

on an arbitrary time interval $[t_a, t_b]$. The behavioural model is derived from the differential invariance which is characteristic of the system and eliminates the need of initial conditions and is in the form of a homogeneous Fredholm integral equation of the second kind with a Hilbert Schmidt kernel [26]. The mathematical interpretation as a Reproducing Kernel Hilbert Space (RKHS) of the behavioural model allows us to extract signal and its time derivatives that confirm the system invariance from output measurement subject to noise. For the OLS estimator to be a BLUE (Best Linear Unbiased Estimator); see [1] the residual has to have constant variance (Homoscedastic).If the error residuals does not have a constant variance,they are termed to be as heteroskedastic.Heteroskedasticity is the most frequent complication in parameter estimation using regression, it has serious consequences for the OLS estimator. Despite that the OLS estimator remains unbiased, the estimated regression error is wrong while confidence intervals cannot be relied on [11]. "A standard quite powerful way to deal with unknown heteroskedasticity is to resort to Feasible Generalised least squares (FGLS), which can be shown to be BLUE (Best Linear Unbiased Estimator)" [12]. The FGLS uses the with inverse covariance weighting recursively replacing the inverse covariance matrix every step in the regression error minimization problem associated with the stochastic multiple regression equations for parameter estimation [2].

1.2 Thesis summary: objectives and organization

"There are two ways of parameter estimation: online and offline. In offline identification, the accumulation of data used to estimate the parameters. All the data are available to be analyzed. On the other hand, online identification is done recursively in time. It means that the parameter estimates are updated recursively within the time limit. In this thesis, the identification using online approach will be used to estimate the parameter within a given model structure" [2].

The Recursive Least Squares method is one of the most popular adaptive filters. As compared to the OLS algorithm, the FGLS offers a superior convergence rate, especially for highly correlated input signals; however, there would be an increase in the computational complexity [17]. The FGLS has been widely used and it has several advantages such as fast parameter convergence and feasible to apply to the direct closed-loop parameter identification approach .

"The tracking ability of the parameter estimator depends on the covariance element but its strength decays with time. To overcome this problem, covariance resetting approach will be employed to revitalize the strength of the tracking variations of parameters. However, it is not easy to employ the resetting technique since the time of parameter change is unknown. One way to detect the changes of parameter is to measure the residual error inside the recursive algorithm. Without the execution of the covariance resetting technique, the residual error will exhibit a sudden change. Covariance resetting technique is executed whenever the algorithm of recursive identification perceives a change. Before the covariance resetting takes place, the residual error will exhibit some irregular movements when the parameter varies. These irregularities are then used to execute covariance resetting algorithm" [2].

Chapter 1 : provides a brief introduction to algebraic parameter and state estimation as the central topic in this thesis. It also states the objectives in the thesis work.

Chapter 2: focuses on the derivation and understanding of the double sided kernels for homogeneous SISO LTI systems and develops the kernel for a fourth order system.

Chapter 3: discusses the previous works done on state and online parameter estimation .

Chapter 4: presents a summary forward-backward kernel based state and parameter estimation using multiple regression equations which is very efficient in the presence of heteroskedasticity .

Chapter 5: is dedicated to numerical results and discussion of a mathematical approach.

Chapter 2

A Double Sided Kernel in SISO LTI Representation [18]

As mentioned in the introduction, algebraic estimation has different advantages where the framework is first laid in [9] and [8]. The theory for the problem considered in this thesis was initially briefly presented in [13]. It employs forward and backward integration and Cauchy formula for multiple integrals to convert a high order differential equation, that embodies a system invariant, into an integral form with no singularities at the boundaries of the observation window. In any observation window, the output reproducing form of system invariance can then be used to characterize system trajectories. An equivalent system representation takes the form of a subspace of an RKHS Hilbert space - a fact that can be used in denoisification of measured output [16].The paper considers only single-input single-output systems and it states the problem as follows.

2.1 Algebraic parameter estimation using kernel representation of homogeneous SISO LTI systems [13]

The estimation problem assumes a SISO LTI system structure :

$$\begin{aligned} \dot{x} &= Ax \\ y &= Cx \end{aligned} \tag{2.1}$$

where the system matrix A is in canonical form,

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-1} \end{bmatrix} \tag{2.2}$$

and,

$$C = \begin{bmatrix} 1 & 0 & 0 & \dots \end{bmatrix} \tag{2.3}$$

so that the characteristic equation of the system is,

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{2.4}$$

The unknown values of the parameters $a_i, i = 0, .., n-1$ need to identified using noisy observations of the system's output $y(t)$ over a finite, but arbitrary, interval of time $t \in [0, T]$, $T > 0$. The input-output equation of the system can be represented as;

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) = -b_{n-1}u^{(n-1)}(t) - \cdots - b_0 u(t) \tag{2.5}$$

In the light of the assumptions made, the system is differentially flat, so that the entire state vector $x(t)$, for $t \in [a, b]$, can be instantaneously recovered from the knowledge of the input and output functions $u(t), y(t), t \in [a, b]$ and their derivatives

2.1.1 A Differential invariant and its controlled version

The LHS and RHS of equation (2.5) involve differential operators F and U acting on the system so that it can be represented as;

$$F(y)(t) = U(u)(t); t \in [a, b] \tag{2.6}$$

If input to the system is not changing for all times then F represents a differential invariant whose output remains constant under the action of the flow of the system. If input u varies in time the equality delivers what is referred as *controlled invariance.*

2.1.2 Kernel Representation [5]

Applying control invariance of (2.6), which in this case coincides with an input-output system equation, an initial condition free integral representation of the system can be obtained. The construction relies on the introduction of functions, which act as annihilators of initial or boundary conditions.

Definition of Annihilators
A pair of smooth (class $\mathcal{C}^{\infty}$) functions

$$(\alpha_a, \alpha_b)\,, \alpha_s : [a, b] \rightarrow \mathbb{R}$$

s = a or b , is an annihiltor of the boundary conditions for system (2.5) if the α_s are non-negative, monotonic and vanish with their derivatives up to order n - 1 at the respective ends of the the interval [a,b]; i.e.

$$\alpha_s^{(i)}(s) = 0 \quad i = 0, \cdots, n-1; \quad s = a, b; \quad \alpha_s^{(0)} \equiv \alpha_s$$

such that their sum is strictly positive, i.e. for some constant c>0,

$$\alpha_{ab}(t) := \alpha_a(t) + \alpha_b(t) > ct \in [a, b]$$

An example for such an annihilator for system (2.5) is the pair

$$\begin{aligned}
&\alpha_a(t) := (t-a)^n, \quad \alpha_b(t) := (b-t)^n \quad t \in [a, b] \\
&\alpha_{ab}(t) := \alpha_a(t) + \alpha_b(t) > 0 \quad t \in [a, b] \\
&\alpha_{ab}(s) = (b-a)^n, \quad s = a, b
\end{aligned}$$

Applying the annihilators the integral representation of the system (2.5) can be obtained as stated below in Theorem 1.
The paper [13] proceeds to develop the kernel representation for a third order system.
In this thesis the double sided kernel approach is presented explicitly for a fourth order

system. The knowledge of the system characteristic equation enables one to replace the state equations with an output reproducing property on an arbitrary time interval $[a, b]$. The differential invariance, which is essentially the characteristic equation, is used to derive the behavioral model of the system [14]. The mathematical interpretation as a Reproducing Kernel Hilbert Space (RKHS) of the behavioral model allows us to extract signal and its time derivatives that confirm the system invariance from output measurement subject to noise [25].

The following theorem describing such system representation has been stated and proved.

Theorem 1. *There exist Hilbert-Schmidt kernels K_{DS}, K^i_{DS}, $i = 1, \cdots n-1$, such that the output function y of* (2.5) *is reproduced on any given interval* $[a, b]$ *in accordance with the action of the evaluation functional*

$$y(t) = \int_a^b K_{DS}(t,\tau)y(\tau)\ \mathrm{d}\tau\ ;\ \forall t \in [a,b] \tag{2.7}$$

and the derivatives of the output $y^{(1)}, \cdots y^{(n-1)}$ can be computed recursively by way of output integration, so that for $i = 1, \cdots n-1$ and for all $t \in [a, b]$:

$$y^{(i)}(t) = \sum_{k=0}^{i-1} b_k(t)y^{(k)}(t) + \int_a^b K^i_{DS}(t,\tau)y(\tau)\ \mathrm{d}\tau \tag{2.8}$$

where $y^{(0)} \equiv y$ and $b_k(\cdot)$ are rational functions of t.Hilbert-Schmidt kernels are square integrable functions on $L^2[a, b] \times L^2[a, b]$.

The kernels of Theorem 1 have the following expression for a system of order n,

$$\begin{aligned} K_{F,y}(n,t,\tau) = &\sum_{j=1}^{n}(-1)^{j+1}\begin{pmatrix} n \\ j \end{pmatrix}\frac{n!(t-\tau)^{j-1}(\tau-a)^{n-j}}{(n-j)!(j-1)!} \\ &+\sum_{i=0}^{n-1}a_i\sum_{j=0}^{i}(-1)^{j+1}\begin{pmatrix} i \\ j \end{pmatrix}\frac{n!(t-\tau)^{n-i+j-1}(\tau-a)^{n-j}}{(n-j)!(n-i+j-1)!} \end{aligned} \tag{2.9}$$

$$K_{B,y}(n,t,\tau) = \sum_{j=1}^{n} \begin{pmatrix} n \\ j \end{pmatrix} \frac{n!(t-\tau)^{j-1}(b-\tau)^{n-j}}{(n-j)!(j-1)!} + \sum_{i=0}^{n-1} a_i \sum_{j=0}^{i} \begin{pmatrix} i \\ j \end{pmatrix} \frac{n!(t-\tau)^{n-i+j-1}(b-\tau)^{n-j}}{(n-j)!(n-i+j-1)!} \tag{2.10}$$

$$K_{F,u}(n,t,\tau) = \sum_{i=0}^{n-1} b_i \sum_{j=0}^{i} (-1)^{j+1} \begin{pmatrix} i \\ j \end{pmatrix} \frac{n!(t-\tau)^{n-i+j-1}(\tau-a)^{n-j}}{(n-j)!(n-i+j-1)!} \tag{2.11}$$

$$K_{B,u}(n,t,\tau) = \sum_{i=0}^{n-1} b_i \sum_{j=0}^{i} \begin{pmatrix} i \\ j \end{pmatrix} \frac{n!(t-\tau)^{n-i+j-1}(b-\tau)^{n-j}}{(n-j)!(n-i+j-1)!} \tag{2.12}$$

2.2 Kernel development for a fourth order system [18]

The general characteristic equation of a fourth order system is written as below:

$$y^{(4)}(t) + a_3 y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{2.13}$$

on an interval $[a,b]$.

Multiplying equation (2.13)by $(\epsilon - a)^4$ and by $(b-\zeta)^4$ we obtain (2.14) and (2.15) respectively.

$$(\varepsilon-a)^4 y^{(4)}(t) + a_3(\varepsilon-a)^4 y^{(3)}(t) + a_2(\varepsilon-a)^4 y^{(2)}(t)_+ a_1(\varepsilon-a)^4 y^{(1)}(t) + a_0(\varepsilon-a)^4 y(t) = 0 \tag{2.14}$$

$$(b-\zeta)^4 y^{(4)}(t) + a_3(b-\zeta)^4 y^{(3)}(t) + a_2(b-\zeta)^4 y^{(2)}(t) + a_1(b-\zeta)^4 y^{(1)}(t) + a_0(b-\zeta)^4 y(t) = 0 \tag{2.15}$$

We will integrate (2.14) and (2.15) four times on the interval $[a, a+\tau]$ and $[b-\sigma, b]$, which essentially means that the (2.13) will be integrated in the forward direction for the interval

$[a, a+\tau]$ and in the backward direction during the interval $[b, b-\sigma]$.
Each term in (2.14) will be separately integrated four times as below,

$$\begin{aligned}
&\int_a^{a+\tau} (\varepsilon - a)^4 \left(y^{(4)}\right)(\varepsilon) d\varepsilon \\
&\quad = \tau^4 y^{(3)}(a+\tau) - \left[4(\varepsilon-a)^3 y^{(2)}\ (\varepsilon)|_a^{a+\tau} - \int_a^{a+\tau} 12(\varepsilon-a)^2 y^{(2)}(\varepsilon) d(\varepsilon) \right] \\
&\quad = \tau^4 y^{(3)}(a+\tau) - 4\tau^3 y^{(2)}(a+\tau) + 12(\varepsilon-a)^2 y\ (\varepsilon)|_a^{a+\tau}\ - \int_a^{a+\tau} 24(\varepsilon-a) y(\varepsilon) d(\varepsilon) \\
&\quad = \tau^4 y^{(3)}(a+\tau) - 4\tau^3 y^{(2)}(a+\tau) + 12\tau^2 y^{(1)}(a+\tau) - 24(\varepsilon-a) y\ (\varepsilon)|_a^{a+\tau} \\
&\qquad\qquad + \int_a^{a+\tau} 24(\varepsilon-a) y(\varepsilon) d(\varepsilon) \qquad (2.16)
\end{aligned}$$

$$\begin{aligned}
\int_a^{a+\tau} (\varepsilon-a)^4 y^{(4)}(\varepsilon) d\varepsilon &= \tau^4 y^{(3)}(a+\tau) - 4\tau^3 y^{(2)}(a+\tau) + 12\tau^2 y^{(1)}(a+\tau) \\
&\quad - 24\tau y(a+\tau) + \int_a^{a+\tau} 24 y(\varepsilon) d\varepsilon
\end{aligned} \qquad (2.17)$$

The upper limit of the integral is made a " dummy variable", in order to integrate the first term of (2.14) again, that is set $\varepsilon' = a + \tau$ and hence,
$\tau^4 y^{(3)}(a+\tau) = (\varepsilon' - a)^4\, y^{(3)}\,(\varepsilon')$
$\tau^3 y^{(2)}(a+\tau) = (\varepsilon' - a)^3\, y^{(2)}\,(\varepsilon')$
$\tau^2 y^{(1)}(a+\tau) = (\varepsilon' - a)^2\, y^{(1)}\,(\varepsilon')$
$\tau y(a+\tau) = (\varepsilon' - a)\, y\,(\varepsilon')$, the integration proceeds with the above changes and integrating (2.14) again,

$$\begin{aligned}
&\int_a^{a+\tau}\int_a^{\varepsilon'}(\varepsilon-a)^4 y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' \\
&\qquad=\int_a^{a+\tau}\left(\varepsilon'-a\right)^4 y^{(3)}\left(\varepsilon'\right)d\varepsilon'-4\int_a^{a+\varepsilon}\left(\varepsilon'-a\right)^3 y^{(2)}\left(\varepsilon'\right)d\varepsilon' \\
&\qquad\quad+12\int_a^{a+\varepsilon}\left(\varepsilon'-a\right)^2 y^{(1)}\left(\varepsilon'\right)d\varepsilon'-24\int_a^{a+\varepsilon}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon \\
&\qquad\quad+24\int_a^{a+\varepsilon}\int_a^{\varepsilon'}y\left(\varepsilon'\right)d\varepsilon d\varepsilon'
\end{aligned}\tag{2.18}$$

$$\begin{aligned}
&\int_a^{a+\tau}\int_a^{\varepsilon'}(\varepsilon-a)^4 y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' \\
&\quad=\left(\varepsilon'-a\right)^4 y^{(2)}\left(\varepsilon'\right)\Big|_a^{a+\tau}-4\int_a^{a+\tau}\left(\varepsilon'-a\right)^3 y^{(2)}\left(\varepsilon'\right)d\varepsilon'-4\left(\varepsilon'-a\right)^3 y\left(\varepsilon'\right)\Big|_a^{a+\tau} \\
&\quad+12\int_a^{a+\tau}\left(\varepsilon'-a\right)^2 y\left(\varepsilon'\right)d\varepsilon'+12\left(\varepsilon'-a\right)^2 y\left(\varepsilon'\right)\Big|_a^{a+\tau} \\
&\quad-24\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon'+24\int_a^{a+\varepsilon}\int_a^{\varepsilon'}y(\varepsilon)d(\varepsilon)d\varepsilon' \\
&\quad-24\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon'
\end{aligned}\tag{2.19}$$

$$\begin{aligned}
&\int_a^{a+\tau}\int_a^{\varepsilon'}(\varepsilon-a)^4 y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' \\
&\quad=\tau^4 y^{(2)}(a+\tau)-4\left(\varepsilon'-a\right)^3 y\ \varepsilon'\Big|_a^{a+\tau}+12\int_a^{a+\tau}\left(\varepsilon'-a\right)^2 y\left(\varepsilon'\right)d\varepsilon' \\
&\quad-4\tau^3 y(a+\tau)+12\left(\varepsilon'-a\right)^2 y\left(\varepsilon'\right)\Big|_a^{a+\tau}-24\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon' \\
&\quad+12\tau^2 y(a+\tau)+24\int_a^{a+\varepsilon}\int_a^{\varepsilon'}y(\varepsilon)d(\varepsilon)d\varepsilon'-48\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon'
\end{aligned}\tag{2.20}$$

$$\begin{aligned}\int_a^{a+\tau}\int_a^{\varepsilon'}(\varepsilon-a)^4y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' \\ =&\tau^4y^{(2)}(a+\tau)-4\tau^3y(a+\tau)+12\left(\varepsilon'-a\right)^2y\left(\varepsilon\right)\big|_a^{a+\tau} \\ &-24\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon'-4\tau^3y(a+\tau)+24\tau^2y(a+\tau) \\ &+24\int_a^{a+\varepsilon}\int_a^{\varepsilon'}y(\varepsilon)d(\varepsilon)d\varepsilon'-72\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon' \end{aligned} \tag{2.21}$$

Finally we get ,

$$\begin{aligned}\int_a^{a+\tau}\int_a^{\varepsilon'}(\varepsilon-a)^4y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' =&\tau^4y^{(2)}(a+\tau)-8\tau^3y(a+\tau)+36\tau^2y(a+\tau) \\ &-96\int_a^{a+\tau}\left(\varepsilon'-a\right)y\left(\varepsilon'\right)d\varepsilon'+24\int_a^{a+\tau}\int_a^{\varepsilon'}y(\varepsilon)d\varepsilon d\varepsilon' \end{aligned} \tag{2.22}$$

To integrate for the third time, we again set $\varepsilon''=a+\tau$ in (2.22) we get,

$$\begin{aligned}\int_a^{a+\tau}\int_a^{\varepsilon'}\int_a^{\varepsilon''}(\varepsilon-a)^4y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' d\varepsilon'' \\ =&\int_a^{a+\tau}\left(\varepsilon''-a\right)^4y^{(2)}\left(\varepsilon''\right)d\varepsilon''-8\int_a^{a+\tau}\left(\varepsilon''-a\right)^3y'\left(\varepsilon''\right)d\varepsilon'' \\ &+36\int_a^{a+\tau}\left(\varepsilon''-a\right)^2y\left(\varepsilon''\right)d\varepsilon''-96\int_a^{a+\tau}\int_a^{\varepsilon''}\left(\varepsilon'-a\right)y\left(\varepsilon\right)d\varepsilon' d\varepsilon'' \\ &+24\int_a^{a+\tau}\int_a^{\varepsilon'}\int_a^{\varepsilon''}y(\varepsilon)d\varepsilon d\varepsilon' d\varepsilon'' \end{aligned} \tag{2.23}$$

$$\begin{aligned}\int_a^{a+\tau}\int_a^{\varepsilon'}\int_a^{\varepsilon''}(\varepsilon-a)^4y^{(4)}(\varepsilon)d\varepsilon d\varepsilon' d\varepsilon'' \\ =&\tau^4y^{(1)}(a+\tau)-4\left(\varepsilon''-a\right)^3y\left(\varepsilon''\right)\big|_a^{a+\tau}+12\int_a^{a+\tau}\left(\varepsilon''-a\right)^2y\left(\varepsilon''\right)d\varepsilon'' \end{aligned}$$

$$
\begin{aligned}
&- 8\tau^3 y(a+\tau) + 60\int_a^{a+\tau} (\varepsilon'' - a)^2\, y(\varepsilon'')\, d\varepsilon'' \\
&- 96\int_a^{a+\tau}\int_a^{\varepsilon''} (\varepsilon' - a)\, y(\varepsilon)\, d\varepsilon' d\varepsilon'' + 24\int_a^{a+\tau}\int_a^{\varepsilon'}\int_a^{\varepsilon''} y(\varepsilon) d\varepsilon d\varepsilon' d\varepsilon''
\end{aligned}
\tag{2.24}
$$

$$
\begin{aligned}
&\int_a^{a+\tau}\int_a^{\varepsilon'}\int_a^{\varepsilon''} (\varepsilon - a)^4 y^{(4)}(\varepsilon) d\varepsilon d\varepsilon' d\varepsilon'' \\
&\quad = \tau^4 y(a+\tau) - 12\tau^3 y(a+\tau) + 72\int_a^{a+\tau} (\varepsilon' - a)^2\, y(\varepsilon'')\, d\varepsilon'' \\
&\quad - 96\int_a^{a+\tau}\int_a^{\varepsilon''} (\varepsilon' - a)\, y(\varepsilon')\, d\varepsilon d\varepsilon'' + 24\int_a^{a+\tau}\int_a^{\varepsilon'}\int_a^{\varepsilon''} y(\varepsilon) d\varepsilon d\varepsilon' d\varepsilon''
\end{aligned}
\tag{2.25}
$$

Replacing the upper limit on the integral by a 'dummy variable', $\varepsilon''' = a+\tau$ and integrating the first term of (2.14) for the fourth time,

$$
\begin{aligned}
&\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} (\varepsilon - a)^4 y^{(4)}(\varepsilon)\, \mathrm{d}\varepsilon \mathrm{d}\varepsilon' \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon''' \\
&\quad = \int_a^{a+\tau} (\varepsilon''' - a)^4 y^{(1)}(\varepsilon''') \mathrm{d}\varepsilon''' - \int_a^{a+\tau} 12(\varepsilon''' - a)^3 y(\varepsilon''') \mathrm{d}\varepsilon''' \\
&\qquad + \int_a^{a+\tau}\int_a^{\varepsilon'''} 72(\varepsilon'' - a)^2 y(\varepsilon'') \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon''' - \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''} 96(\varepsilon' - a) y(\varepsilon') \mathrm{d}\varepsilon' \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon'''
\end{aligned}
\tag{2.26}
$$

$$
\begin{aligned}
&+ \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 24 y(\varepsilon) \mathrm{d}\varepsilon \mathrm{d}\varepsilon' \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon''' \\
&= \tau^4 y(a+\tau) - \int_a^{a+\tau} 16(\varepsilon''' - a)^3 y(\varepsilon''') \mathrm{d}\varepsilon''' + \int_a^{a+\tau}\int_a^{\varepsilon'''} 72(\varepsilon'' - a)^2 y(\varepsilon'') \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon''' \\
&\quad - \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''} 96(\varepsilon' - a) y(\varepsilon') \mathrm{d}\varepsilon' \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon''' + \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 24 y(\varepsilon) \mathrm{d}\varepsilon \mathrm{d}\varepsilon' \mathrm{d}\varepsilon'' \mathrm{d}\varepsilon'''
\end{aligned}
\tag{2.27}
$$

Integrating the second term in (2.14) for the first time,

$$\int_a^{a+\tau} a_3(\varepsilon-a)^4 y^{(3)}(\varepsilon)\mathrm{d}\varepsilon$$

$$= a_3(\varepsilon-a)^4 y^{(2)}(\varepsilon)\mid_a^{a+\tau} - \int_a^{a+\tau} 4a_3(\varepsilon-a)^3 y^{(2)}(\varepsilon)\mathrm{d}\varepsilon$$

$$= a_3\tau^4 y^{(2)}(a+\tau) - \left[4a_3(\varepsilon-a)^3 y^{(1)}(\varepsilon)\mid_a^{a+\tau} - \int_a^{a+\tau} 12a_3(\varepsilon-a)^2 y^{(1)}(\varepsilon)\mathrm{d}\varepsilon\right]$$

$$= a_3\tau^4 y^{(2)}(a+\tau) - 4a_3\tau^3 y^{(1)}(a+\tau) + 12a_3\tau^2 y(a+\tau) - \int_a^{a+\tau} 24a_3(\varepsilon-a)y(\varepsilon)\mathrm{d}\varepsilon \quad (2.28)$$

Following similar steps as before and introducing 'dummy variable', $\varepsilon' = a+\tau$, and integrating for the second time,

$$\int_a^{a+\tau}\int_a^{\varepsilon'} a_3(\varepsilon-a)^4 y^{(3)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'$$

$$= \int_a^{a+\tau} a_3(\varepsilon'-a)^4 y^{(2)}(\varepsilon')\mathrm{d}\varepsilon' - \int_a^{a+\tau} 4a_3(\varepsilon'-a)^3 y^{(1)}(\varepsilon')\mathrm{d}\varepsilon' + \int_a^{a+\tau} 12a_3(\varepsilon'-a)^2 y(\varepsilon')\mathrm{d}\varepsilon'$$

$$- \int_a^{a+\tau}\int_a^{\varepsilon'} 24a_3(\varepsilon-a)y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'$$

$$= a_3\tau^4 y^{(1)}(a+\tau) - 8a_3\tau^3 y(a+\tau) + \int_a^{a+\tau} 36a_3(\varepsilon'-a)^2 y(\varepsilon')\mathrm{d}\varepsilon'$$

$$- \int_a^{a+\tau}\int_a^{\varepsilon'} 24a_3(\varepsilon-a)y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon' \quad (2.29)$$

Integrating for the third time,

$$\int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_3(\varepsilon-a)^4 y^{(3)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$
$$= \int_a^{a+\tau} a_3(\varepsilon''-a)^4 y^{(1)}(\varepsilon'')\mathrm{d}\varepsilon'' - \int_a^{a+\tau} 8a_3(\varepsilon''-a)^3 y(\varepsilon'')\mathrm{d}\varepsilon''$$
$$+ \int_a^{a+\tau}\int_a^{\varepsilon''} 36a_3(\varepsilon'-a)^2 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon'' - \int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 24a_3(\varepsilon-a)y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$
$$= a_3\tau^4 y(a+\tau) - \int_a^{a+\tau} 12a_3(\varepsilon''-a)^3 y(\varepsilon'')\mathrm{d}\varepsilon'' + \int_a^{a+\tau}\int_a^{\varepsilon''} 36a_3(\varepsilon'-a)^2 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$
$$- \int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 24a_3(\varepsilon-a)y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon'' \tag{2.30}$$

Repeating the steps and integrating the second term in (2.14) for the fourth time,

$$\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_3(\varepsilon-a)^4 y^{(3)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$
$$= \int_a^{a+\tau} a_3(\varepsilon'''-a)^4 y(\varepsilon''')\mathrm{d}\varepsilon''' - \int_a^{a+\tau}\int_a^{\varepsilon'''} 12a_3(\varepsilon''-a)^3 y(\varepsilon'')\mathrm{d}\varepsilon''\mathrm{d}\varepsilon'''$$
$$+ \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''} 36a_3(\varepsilon'-a)^2 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' - \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 24a_3(\varepsilon-a)y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' \tag{2.31}$$

Integrating the third term in (2.14) once,

$$\int_a^{a+\tau} a_2(\varepsilon-a)^4 y^{(2)}(\varepsilon)\mathrm{d}\varepsilon = a_2(\varepsilon-a)^4 y^{(1)}(\varepsilon)\,\Big|_a^{a+\tau} - \int_a^{a+\tau} 4a_2(\varepsilon-a)^3 y^{(1)}(\varepsilon)\mathrm{d}\varepsilon$$

$$= a_2\tau^4 y^{(1)}(a+\tau) - \left[4a_2(\varepsilon - a)^3 y(\varepsilon) \mid_a^{a+\tau} - \int_a^{a+\tau} 12a_2(\varepsilon - a)^2 y(\varepsilon)\mathrm{d}\varepsilon\right]$$

$$= a_2\tau^4 y^{(1)}(a+\tau) - 4a_2\tau^3 y(a+\tau) + \int_a^{a+\tau} 12a_2(\varepsilon - a)^2 y(\varepsilon)\mathrm{d}\varepsilon \tag{2.32}$$

Integrating for the second time,

$$\int_a^{a+\tau}\int_a^{\varepsilon'} a_2(\varepsilon - a)^4 y^{(2)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'$$

$$= \int_a^{a+\tau} a_2(\varepsilon' - a)^4 y^{(1)}(\varepsilon')\mathrm{d}\varepsilon' - \int_a^{a+\tau} 4a_2(\varepsilon' - a)^3 y(\varepsilon')\mathrm{d}\varepsilon' + \int_a^{a+\tau}\int_a^{\varepsilon'} 12a_2(\varepsilon - a)^2 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'$$

$$= a_2\tau^4 y(a+\tau) - \int_a^{a+\tau} 8a_2(\varepsilon' - a)^3 y(\varepsilon')\mathrm{d}\varepsilon' + \int_a^{a+\tau}\int_a^{\varepsilon'} 12a_2(\varepsilon - a)^2 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon' \tag{2.33}$$

Integrating the third term for the third time,

$$\int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_2(\varepsilon - a)^4 y^{(2)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$

$$= \int_a^{a+\tau} a_2(\varepsilon'' - a)^4 y(\varepsilon'')\mathrm{d}\varepsilon'' - \int_a^{a+\tau}\int_a^{\varepsilon''} 8a_2(\varepsilon' - a)^3 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$

$$+ \int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 12a_2(\varepsilon - a)^2 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon'' \tag{2.34}$$

Integrating the above equation again,

$$\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_2(\varepsilon - a)^4 y^{(2)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon'''$$

$$= \int_a^{a+\tau}\int_a^{\varepsilon'''} a_2(\varepsilon''-a)^4 y(\varepsilon'')\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' - \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''} 8a_2(\varepsilon'-a)^3 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon'''$$
$$+ \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 12a_2(\varepsilon-a)^2 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' \tag{2.35}$$

Integrating the fourth term in (2.14),

$$\int_a^{a+\tau} a_1(\varepsilon-a)y^{(1)}(\varepsilon)\mathrm{d}\varepsilon$$
$$= a_1(\varepsilon-a)^4 y(\varepsilon)\mid_a^{a+\tau} - \int_a^{a+\tau} 4a_1(\varepsilon-a)^3 y(\varepsilon)\mathrm{d}\varepsilon$$
$$= a_1\tau^4 y(a+\tau) - \int_a^{a+\tau} 4a_1(\varepsilon-a)^3 y(\varepsilon)\mathrm{d}\varepsilon \tag{2.36}$$

By introducing the dummy variable and integrating for the second time,

$$\int_a^{a+\tau}\int_a^{\varepsilon'} a_1(\varepsilon-a)^4 y^{(1)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'$$
$$= \int_a^{a+\tau} a_1(\varepsilon'-a)^4 y(\varepsilon')\mathrm{d}\varepsilon' - \int_a^{a+\tau}\int_a^{\varepsilon'} 4a_1(\varepsilon-a)^3 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon' \tag{2.37}$$

Integrating the above equation again,

$$\int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_1(\varepsilon-a)^4 y^{(1)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon'' = \int_a^{a+\tau}\int_a^{\varepsilon''} a_1(\varepsilon'-a)^4 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''$$
$$- \int_a^{a+\tau}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 4a_1(\varepsilon-a)^3 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon'' \tag{2.38}$$

Finally integrating the fourth term of (2.14) for the fourth time,

$$\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_1(\varepsilon-a)^4 y^{(1)}(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon'''$$
$$=\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''} a_1(\varepsilon'-a)^4 y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' - \int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} 4a_1(\varepsilon-a)^3 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' \tag{2.39}$$

Integrating the last term in (2.14) four times we get,

$$\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'} a_0(\varepsilon-a)^4 y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' \tag{2.40}$$

Collecting all the terms after being subjected to integration four times , we get,

$$\tau^4 y(a+\tau) = \int_a^{a+\tau}\Big[16(\varepsilon'''-a)^3 - a_3(\varepsilon'''-a)^4\Big]y(\varepsilon''')\mathrm{d}\varepsilon'''$$
$$+\int_a^{a+\tau}\int_a^{\varepsilon'''}\Big[-72(\varepsilon''-a)^2 + 12a_3(\varepsilon''-a)^3 - a_2(\varepsilon''-a)^4\Big]y(\varepsilon'')\mathrm{d}\varepsilon''\mathrm{d}\varepsilon'''$$
$$+\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\Big[96(\varepsilon'-a) - 36a_3(\varepsilon'-a)^2 + 8a_2(\varepsilon'-a)^3 - a_1(\varepsilon'-a)^4\Big]y(\varepsilon')\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon'''$$
$$+\int_a^{a+\tau}\int_a^{\varepsilon'''}\int_a^{\varepsilon''}\int_a^{\varepsilon'}\Big[-24 + 24a_3(\varepsilon-a) - 12a_2(\varepsilon-a)^2 +$$
$$4a_1(\varepsilon-a)^3 - a_0(\varepsilon-a)^4\Big]y(\varepsilon)\mathrm{d}\varepsilon\mathrm{d}\varepsilon'\mathrm{d}\varepsilon''\mathrm{d}\varepsilon''' \tag{2.41}$$

Cauchy's formula for repeated integration (stated below) is used in order to simplify the above equation and is explained below.

Let f be a continuous function on the real line, then the nth repeated integral of f based at a,

$$f^{(-n)}(x) = \int_a^x \int_a^{\sigma_1} \cdots \int_a^{\sigma_{n-1}} f(\sigma_n) \mathrm{d}\sigma_n \cdots \mathrm{d}\sigma_2 \mathrm{d}\sigma_1 \tag{2.42}$$

is equivalent to a single integration

$$f^{(-n)}(x) = \frac{1}{(n-1)!} \int_a^x (x-t)^{n-1} f(t) \mathrm{d}t \tag{2.43}$$

We let $a + \tau = t$ in (2.41) and apply Cauchy's formula,

$$(t-a)^4 y(t) \triangleq \int\limits_a^t K_{F,y}(t,\tau) y(\tau)\, \mathrm{d}\tau \tag{2.44}$$

with $K_{F,y}(t,\tau)$ as,

$$\begin{aligned} K_{F,y}(t,\tau) = &\Big[16(\tau-a)^3 - a_3(\tau-a)^4\Big] \\ &+ (t-\tau)\Big[-72(\tau-a)^2 + 12a_3(\tau-a)^3 - a_2(\tau-a)^4\Big] \\ &+ \frac{(t-\tau)^2}{2}\Big[96(\tau-a) - 36a_3(\tau-a)^2 + 8a_2(\tau-a)^3 - a_1(\tau-a)^4\Big] \\ &+ \frac{(t-\tau)^3}{6}\Big[-24 + 24a_3(\tau-a) - 12a_2(\tau-a)^2 + 4a_1(\tau-a)^3 - a_0(\tau-a)^4\Big] \end{aligned} \tag{2.45}$$

Now consider the equation (2.15),

$$(b-\zeta)^4 y^{(4)}(t) + a_3(b-\zeta)^4 y^{(3)}(t) + a_2(b-\zeta)^4 y^{(2)}(t) + a_1(b-\zeta)^4 y^{(1)}(t) + a_0(b-\zeta)^4 y(t) = 0$$

$$\int_{b-\sigma}^{b} (b-\zeta)^4 y^{(4)}(\zeta)\mathrm{d}\zeta$$

$$= (b-\zeta)^4 y^{(3)}\zeta \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 4(b-\zeta)^3 y^{(3)}(\zeta)\mathrm{d}\zeta$$

$$= -\sigma^4 y^{(3)}(b-\sigma) + \left[4(b-\zeta)^3 y^{(2)}(\zeta) \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 12(b-\zeta)^2 y^{(2)}(\zeta)\mathrm{d}\zeta \right]$$

$$= -\sigma^4 y^{(3)}(b-\sigma) - 4\sigma^3 y^{(2)}(b-\sigma) + \left[12(b-\zeta)^2 y^{(1)}(\zeta) \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 24(b-\zeta) y^{(1)}(\zeta)\mathrm{d}\zeta \right]$$

$$= -\sigma^4 y^{(3)}(b-\sigma) - 4\sigma^3 y^{(2)}(b-\sigma) - 12\sigma^2 y^{(1)}(b-\sigma) + \left[24(b-\zeta) y(\zeta) \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 24 y(\zeta)\mathrm{d}\zeta \right]$$

$$= -\sigma^4 y^{(3)}(b-\sigma) - 4\sigma^3 y^{(2)}(b-\sigma) - 12\sigma^2 y^{(1)}(b-\sigma) - 24\sigma y(b-\sigma) + \int_{b-\sigma}^{b} 24 y(\zeta)\mathrm{d}\zeta \tag{2.46}$$

As shown in the forward integration procedure earlier, the upper limit on the integral is replaced by a 'dummy variable ', $\zeta' = b - \sigma$, meaning

$$-\sigma^4 y^{(3)}(b-\sigma) = -(b-\zeta')^4 y^{(3)}(\zeta')$$

$$-4\sigma^3 y^{(2)}(b-\sigma) = -4(b-\zeta')^3 y^{(2)}(\zeta')$$

$$-12\sigma^2 y^{(1)}(b-\sigma) = -12(b-\zeta')^2 y^{(1)}(\zeta')$$

$$-24\sigma y(b-\sigma) = -24(b-\zeta') y(\zeta')$$

Now, we integrate the first term of (2.15) for the second time,

$$\int_{b-\sigma}^{b}\int_{\zeta'}^{b}(b-\zeta)^4 y^{(4)}(\zeta)\mathrm{d}\zeta$$

$$= -\int_{b-\sigma}^{b}(b-\zeta)^4 y^{(3)}(\zeta')\mathrm{d}\zeta' - \int_{b-\sigma}^{b}4(b-\zeta')^3 y^{(2)}(\zeta')\mathrm{d}\zeta'$$

$$-\int_{b-\sigma}^{b}12(b-\zeta')^2 y^{(1)}(\zeta')\mathrm{d}\zeta' - \int_{b-\sigma}^{b}24(b-\zeta')y(\zeta')\mathrm{d}\zeta' + \int_{b-\sigma}^{b}\int_{\zeta'}^{b}24y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'$$

$$= \sigma^4 y^{(2)}(b-\sigma) + 8\sigma^3 y^{(1)}(b-\sigma) + 36\sigma^2 y(b-\sigma) - \int_{b-\sigma}^{b}96(b-\zeta')y(\zeta')\mathrm{d}\zeta'$$

$$+\int_{b-\sigma}^{b}\int_{\zeta'}^{b}24y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta' \qquad (2.47)$$

Replace the upper limit on the integral by a 'dummy variable', $\zeta'' = b - \sigma$ and integrating the first term of (2.15) for the third time,

$$\int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b}(b-\zeta)^4 y^{(4)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''$$

$$= \int_{b-\sigma}^{b}(b-\zeta'')^4 y^{(2)}(\zeta'')\mathrm{d}\zeta'' + \int_{b-\sigma}^{b}8(b-\zeta'')^3 y^{(1)}(\zeta'')\mathrm{d}\zeta'' + \int_{b-\sigma}^{b}36(b-\zeta'')^2 y(\zeta'')\mathrm{d}\zeta''$$

$$-\int_{b-\sigma}^{b}\int_{\zeta''}^{b}96(b-\zeta')y(\zeta')\mathrm{d}\zeta' + \int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b}24y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''$$

$$= -\sigma^4 y^{(1)}(b-\sigma) - 12\sigma^3 y(b-\sigma) + \int_{b-\sigma}^{b}72(b-\zeta'')^2 y(\zeta'')\mathrm{d}\zeta''$$

$$- \int_{b-\sigma}^{b} \int_{\zeta''}^{b} 96(b-\zeta')y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta'' + \int_{b-\sigma}^{b} \int_{\zeta''}^{b} \int_{\zeta'}^{b} 24y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta'' \tag{2.48}$$

Replace the upper limit on the integral by a 'dummy variable', $\zeta''' = b - \sigma$, and integrating for the fourth time,

$$\begin{aligned}
&\int_{b-\sigma}^{b} \int_{\zeta'''}^{b} \int_{\zeta''}^{b} \int_{\zeta'}^{b} (b-\zeta)^4 y^{(4)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \\
&\quad = - \int_{b-\sigma}^{b} (b-\zeta''')^4 y^{(1)}(\zeta''')\mathrm{d}\zeta''' - \int_{b-\sigma}^{b} 12(b-\zeta''')^3 y(\zeta''')\mathrm{d}\zeta''' \\
&\qquad + \int_{b-\sigma}^{b} \int_{\zeta'''}^{b} 72(b-\zeta'')^2 y(\zeta'')\mathrm{d}\zeta''\zeta''' - \int_{b-\sigma}^{b} \int_{\zeta'''}^{b} \int_{\zeta''}^{b} 96(b-\zeta')y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \\
&\qquad + \int_{b-\sigma}^{b} \int_{\zeta'''}^{b} \int_{\zeta''}^{b} \int_{\zeta'}^{b} 24y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''
\end{aligned} \tag{2.49}$$

Integrating the second term in (2.15) for the first time, we get,

$$\begin{aligned}
&\int_{b-\sigma}^{b} a_3(b-\zeta)^4 y^{(3)}(\zeta)\mathrm{d}\zeta \\
&\quad = a_3(b-\zeta)^4 y^{(2)}(\zeta) \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 4a_3(b-\zeta)^2 y^{(2)}(\zeta)\mathrm{d}\zeta \\
&\quad = -a_3\sigma^4 y^{(2)}(b-\sigma) + \left[4a_3(b-\zeta)y^{(1)}(\zeta) \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 12a_3(b-\zeta)^2 y^{(1)}(\zeta)\mathrm{d}\zeta \right] \\
&\quad = -a_3\sigma^4 y^{(2)}(b-\sigma) - 4a_3\sigma^3 y^{(1)}(b-\sigma) - 12a_3\sigma^2 y(b-\sigma) \\
&\qquad\qquad + \int_{b-\sigma}^{b} 24a_3(b-\zeta)y(\zeta)\mathrm{d}\zeta
\end{aligned} \tag{2.50}$$

Introducing the 'dummy variable', $\zeta' = b - \sigma$ and integrating the above equation again,

$$\begin{aligned}
&\int_{b-\sigma}^{b}\int_{\zeta'}^{b} a_3(b-\zeta)^4 y^3(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta' \\
&\qquad = -\int_{b-\sigma}^{b} a_3(b-\zeta')^4 y^{(2)}(\zeta')\mathrm{d}\zeta' - \int_{b-\sigma}^{b} 4a_3(b-\zeta')^3 y^{(1)}(\zeta')\mathrm{d}\zeta' \\
&\qquad\quad - \int_{b-\sigma}^{b} 12a_3(b-\zeta')^2 y(\zeta')\mathrm{d}\zeta' + \int_{b-\sigma}^{b}\int_{\zeta'}^{b} 24a_3(b-\zeta)y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta' \\
&\qquad = a_3\sigma^4 y^{(1)}(b-\sigma) + 8a_3\sigma^{(3)} y(b-\sigma) - \int_{b-\sigma}^{b} 36a_3(b-\zeta')^2 y(\zeta')\mathrm{d}\zeta' \\
&\qquad\quad + \int_{b-\sigma}^{b}\int_{\zeta'}^{b} 24a_3(b-\zeta)y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'
\end{aligned} \tag{2.51}$$

Integrating the above again , meaning integrating the third term in (2.15) for the third time:

$$\begin{aligned}
&\int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} a_3(b-\zeta)^4 y^3(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta'' \\
&= \int_{b-\sigma}^{b} a_3(b-\zeta'')^4 y^{(1)}(\zeta'')\mathrm{d}\zeta'' + \int_{b-\sigma}^{b} 8a_3(b-\zeta'')^3 y(\zeta'')\mathrm{d}\zeta'' \\
&\quad - \int_{b-\sigma}^{b}\int_{\zeta''}^{b} 36a_3(b-\zeta')^2 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta'' + \int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} 24a_3(b-\zeta)y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta'' \\
&= -a_3\sigma^4 y(b-\sigma) + \int_{b-\sigma}^{b} 12a_3(b-\zeta'')^3 y(\zeta'')\mathrm{d}\zeta'' - \int_{b-\sigma}^{b}\int_{\zeta''}^{b} 36a_3(b-\zeta')^2 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta'' \\
&\quad + \int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} 24a_3(b-\zeta)y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''
\end{aligned} \tag{2.52}$$

Integrating again,

$$\int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} a_3(b-\zeta)^4 y^3(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''$$

$$= -\int_{b-\sigma}^{b} a_3(b-\zeta''')^4 y(\zeta''')\mathrm{d}\zeta''' + \int_{b-\sigma}^{b}\int_{\zeta'''}^{b} 12a_3(b-\zeta'')^3 y(\zeta'')\mathrm{d}\zeta''\mathrm{d}\zeta'''$$

$$-\int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b} 36a_3(b-\zeta')^2 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''$$

$$+\int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} 24a_3(b-\zeta)y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta''\mathrm{d}\zeta''\mathrm{d}\zeta''' \tag{2.53}$$

The third term of equation (2.15) is integrated once, and we get,

$$\int_{b-\sigma}^{b} a_2(b-\zeta)^4 y^{(2)}(\zeta)\mathrm{d}\zeta$$

$$= a_2(b-\zeta)^4 y^{(1)}(\zeta)\,\Big|_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 4a_2(b-\zeta)^3 y^{(1)}(\zeta)\mathrm{d}\zeta$$

$$= -a_2\sigma^4 y^{(1)}(b-\sigma) + \left[4a_2(b-\zeta)^3 y(\zeta)\,\Big|_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 12a_2(b-\zeta)^2 y(\zeta)\mathrm{d}\zeta\right]$$

$$= -a_2\sigma^4 y^{(1)}(b-\sigma) - 4a_2\sigma^3 y(b-\sigma) + \int_{b-\sigma}^{b} 12a_2(b-\zeta)^2 y(\zeta)\mathrm{d}\zeta \tag{2.54}$$

Integrating for the second time,

$$\int_{b-\sigma}^{b}\int_{\zeta'}^{b} a_2(b-\zeta')^4 y^{(2)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta' = -\int_{b-\sigma}^{b} a_2(b-\zeta')^4 y^{(1)}(\zeta')$$

$$
- \int_{b-\sigma}^{b} 4a_2(b-\zeta')^3 y(\zeta')\mathrm{d}\zeta' + \int_{b-\sigma}^{b}\int_{\zeta'}^{b} 12a_2(b-\zeta')^2 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'
$$

$$
= a_2\sigma^4 y(b-\sigma) - \int_{b-\sigma}^{b} 8a_2(b-\zeta')^3 y(\zeta')\mathrm{d}\zeta' + \int_{b-\sigma}^{b}\int_{\zeta'}^{b} 12a_2(b-\zeta)^2 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta' \tag{2.55}
$$

Third time integration yields,

$$
\int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} a_2(b-\zeta)^4 y^{(2)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''
$$

$$
= \int_{b-\sigma}^{b} a_2(b-\zeta'')^{(4)} y(\zeta'')\mathrm{d}\zeta'' - \int_{b-\sigma}^{b}\int_{\zeta''}^{b} 8a_2(b-\zeta')^3 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''
$$

$$
+ \int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} 12a_2(b-\zeta)^2 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta'' \tag{2.56}
$$

Finally, integrating the third term of (2.15) for the fourth time,

$$
\int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} a_2(b-\zeta)^4 y^{(2)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''
$$

$$
= \int_{b-\sigma}^{b}\int_{\zeta'''}^{b} a_2(b-\zeta'')^4 y(\zeta'')\mathrm{d}\zeta''\mathrm{d}\zeta''' - \int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b} 8a_2(b-\zeta')^3 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''
$$

$$
+ \int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} 12a_2(b-\zeta)^2 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \tag{2.57}
$$

Integrating the fourth term in (2.15) once, we get,

$$
\int_{b-\sigma}^{b} a_1(b-\zeta)^4 y^{(1)}(\zeta)\mathrm{d}\zeta = a_1(b-\zeta)^4 y(\zeta) \mid_{b-\sigma}^{b} + \int_{b-\sigma}^{b} 4a_1(b-\zeta)^3 y(\zeta)\mathrm{d}\zeta
$$

$$= -a_1\sigma^4 y(b-\sigma) + \int\limits_{b-\sigma}^{b} 4a_1(b-\zeta)^3 y(\zeta)\mathrm{d}\zeta \tag{2.58}$$

Integrating for the second time,

$$\int\limits_{b-\sigma}^{b}\int\limits_{\zeta'}^{b} a_1(b-\zeta)^4 y^{(1)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'$$
$$= -\int\limits_{b-\sigma}^{b} a_1(b-\zeta')^4 y(\zeta')\mathrm{d}\zeta' + \int\limits_{b-\sigma}^{b}\int\limits_{\zeta'}^{b} 4a_1(b-\zeta)^3 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta' \tag{2.59}$$

Integrating the above equation once more,

$$\begin{aligned}\int\limits_{b-\sigma}^{b}\int\limits_{\zeta''}^{b}\int\limits_{\zeta'}^{b} a_1(b-\zeta)^4 y^{(1)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta'' &= -\int\limits_{b-\sigma}^{b}\int\limits_{\zeta''}^{b} a_1(b-\zeta')^4 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta'' \\ &\quad + \int\limits_{b-\sigma}^{b}\int\limits_{\zeta''}^{b}\int\limits_{\zeta'}^{b} 4a_1(b-\zeta)^3 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\end{aligned} \tag{2.60}$$

Integrating the fourth term for the fourth time,

$$\begin{aligned}\int\limits_{b-\sigma}^{b}\int\limits_{\zeta'''}^{b}\int\limits_{\zeta''}^{b}\int\limits_{\zeta'}^{b} a_1(b-\zeta)^4 y^{(1)}(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' &= -\int\limits_{b-\sigma}^{b}\int\limits_{\zeta'''}^{b}\int\limits_{\zeta''}^{b} a_1(b-\zeta')^4 y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \\ &\quad + \int\limits_{b-\sigma}^{b}\int\limits_{\zeta'''}^{b}\int\limits_{\zeta''}^{b}\int\limits_{\zeta'}^{b} 4a_1(b-\zeta)^3 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''\end{aligned} \tag{2.61}$$

Repeating the above for the fifth term in (2.15), we get,

$$\int\limits_{b-\sigma}^{b}\int\limits_{\zeta'''}^{b}\int\limits_{\zeta''}^{b}\int\limits_{\zeta'}^{b} a_0(b-\zeta)^4 y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \tag{2.62}$$

Collecting terms from the integrals of individual terms of (2.15) we have,

$$\sigma^4 y(b-\sigma)$$
$$= \int_{b-\sigma}^{b} \Big[16(b-\zeta''')^3 + a_3(b-\zeta''')^4\Big] y(\zeta''')\mathrm{d}\zeta'''$$
$$+ \int_{b-\sigma}^{b}\int_{\zeta'''}^{b} \Big[-72(b-\zeta'')^2 - 12a_3(b-\zeta'')^3 - a_2(b-\zeta'')^4\Big] y(\zeta'')\zeta''\zeta'''$$
$$+ \int_{b-\sigma}^{b}\int_{\zeta''}^{b}\int_{\zeta''}^{b} \Big[96(b-\zeta') + 36a_3(b-\zeta')^2 + 8a_2(b-\zeta')^3 + a_1(b-\zeta')^4\Big] y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''$$
$$+ \int_{b-\sigma}^{b}\int_{\zeta'''}^{b}\int_{\zeta''}^{b}\int_{\zeta'}^{b} \Big[-24 - 24a_3(b-\zeta) - 12a_2(b-\zeta)^2$$
$$- 4a_1(b-\zeta)^3 - a_0(b-\zeta)^4\Big] y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \tag{2.63}$$

In order to apply Cauchy's formula, all the limits on the integrals in (2.63) are reversed and making the necessary sign changes, we get

$$\sigma^4 y(b-\sigma) = \int_{b}^{b-\sigma} \Big[-16(b-\zeta''')^3 - a_3(b-\zeta''')^4\Big] y(\zeta''')\mathrm{d}\zeta'''$$
$$+ \int_{b}^{b-\sigma}\int_{b}^{\zeta'''} \Big[-72(b-\zeta'')^2 - 12a_3(b-\zeta'')^3 - a_2(b-\zeta'')^4\Big] y(\zeta'')\zeta''\zeta'''$$
$$+ \int_{b}^{b-\sigma}\int_{b}^{\zeta''}\int_{b}^{\zeta''} \Big[-96(b-\zeta') - 36a_3(b-\zeta')^2 - 8a_2(b-\zeta')^3 - a_1(b-\zeta')^4\Big] y(\zeta')\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta'''$$
$$+ \int_{b}^{b-\sigma}\int_{b}^{\zeta'''}\int_{b}^{\zeta''}\int_{b}^{\zeta'} \Big[-24 - 24a_3(b-\zeta) - 12a_2(b-\zeta)^2$$
$$- 4a_1(b-\zeta)^3 - a_0(b-\zeta)^4\Big] y(\zeta)\mathrm{d}\zeta\mathrm{d}\zeta'\mathrm{d}\zeta''\mathrm{d}\zeta''' \tag{2.64}$$

Substituting $b - \sigma = t$ and applying the formula for repeated integrals as listed in (2.42) and (2.43),

$$
\begin{aligned}
(b-t)^4 y(t) = &\int_b^t \Big[-16(b-\sigma)^3 - a_3(b-\sigma)^4 \Big] y(\sigma)\mathrm{d}\sigma \\
&+ \int_b^t (t-\sigma)\Big[-72(b-\sigma)^2 - 12a_3(b-\sigma)^3 - a_2(b-\sigma)^4 \Big] y(\sigma)\mathrm{d}\sigma \\
&+ \int_b^t \frac{(t-\sigma)^2}{2}\Big[-96(b-\sigma) - 36a_3(b-\sigma)^2 - 8a_2(b-\sigma)^3 - a_1(b-\sigma)^4 \Big] y(\sigma)\mathrm{d}\sigma \\
&+ \int_b^t \frac{(t-\sigma)^3}{6}\Big[-24 - 24a_3(b-\sigma) - 12a_2(b-\sigma)^2 \\
&\qquad - 4a_1(b-\sigma)^3 - a_0(b-\sigma)^4 \Big] y(\sigma)\mathrm{d}\sigma
\end{aligned} \tag{2.65}
$$

We reverse the limits of integration again, and change the variable of integration from σ to τ and get,

$$
\begin{aligned}
(b-t)^4 y(t) = &\int_t^b \Big[16(b-\tau)^3 + a_3(b-\tau)^4 \Big] y(\tau)\mathrm{d}\tau \\
&+ \int_t^b (t-\tau)\Big[72(b-\tau)^2 + 12a_3(b-\tau)^3 + a_2(b-\tau)^4 \Big] y(\tau)\mathrm{d}\tau \\
&+ \int_t^b \frac{(t-\tau)^2}{2}\Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4 \Big] y(\tau)\mathrm{d}\tau \\
&+ \int_t^b \frac{(t-\tau)^3}{6}\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2 + 4a_1(b-\tau)^3 \\
&\qquad + a_0(b-\tau)^4 \Big] y(\tau)\mathrm{d}\tau
\end{aligned} \tag{2.66}
$$

Thereby obtaining,

$$(b-t)^4 y(t) \triangleq \int_t^b K_{B,y}(t,\tau) y(\tau)\,\mathrm{d}\tau \tag{2.67}$$

with

$$\begin{aligned} K_{B,y}(t,\tau) = &\Big[16(b-\tau)^3 + a_3(b-\tau)^4\Big] \\ &+ (t-\tau)\Big[72(b-\tau)^2 + 12a_3(b-\tau)^3 + a_2(b-\tau)^4\Big] \\ &+ \frac{(t-\tau)^2}{2}\Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4\Big] \\ &+ \frac{(t-\tau)^3}{6}\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2 + 4a_1(b-\tau)^3 + a_0(b-\tau)^4\Big] \end{aligned} \tag{2.68}$$

Therefore, adding equations (2.44) and (2.67) and dividing both sides by $[(t-a)^4 + (b-t)^4]$ yields:

$$y(t) = \int_a^b K_{DS,y}(t,\tau) y(\tau)\mathrm{d}\tau \tag{2.69}$$

where,

$$K_{DS,y} \triangleq \frac{1}{[(t-a)^4 + (b-t)^4]} \begin{cases} K_{F,y}(t,\tau) : \tau \le t \\ K_{B,y}(t,\tau) : \tau > t \end{cases} \tag{2.70}$$

In order to find the recursive expressions for the derivatives of the output $y_{(t)}$ similar derivation can be used. To obtain the expression for $y^{(1)}(t)$, equations (2.14) and (2.15) need to be integrated three times each. This gives the following expressions:

$$\begin{aligned}
&(t-a)^{(4)}y^{(1)}(t)\\
&=\Big[12(t-a)^3 - a_3(t-a)^4\Big]y(t)\\
&+\int_a^t \Big[-72(\tau-a)^2 + 12a_3(\tau-a)^3 - a_2(\tau-a)^4\Big]y(\tau)\mathrm{d}\tau\\
&+\int_a^t (t-\tau)\Big[96(\tau-a) - 36a_3(\tau-a)^2 + 8a_2(\tau-a)^3 - a_1(\tau-a)^4\Big]y(\tau)\mathrm{d}\tau\\
&+\int_a^t \frac{(t-\tau)^2}{2}\Big[-24 + 24a_3(\tau-a) - 12a_2(\tau-a)^2\\
&\qquad\qquad + 4a_1(\tau-a)^3 - a_0(\tau-a)^4\Big]y(\tau)\mathrm{d}\tau\\
&+\int_a^t \frac{(t-\tau)^2}{2}(\tau-a)^4u(\tau)\mathrm{d}\tau
\end{aligned} \tag{2.71}$$

$$\begin{aligned}
&(b-t)^{(4)}y^{(1)}(t)\\
&=\Big[-12(b-t)^3 - a_3(b-t)^4\Big]y(t)\\
&+\int_t^b \Big[72(b-\tau)^2 + 12a_3(b-\tau)^3 + a_2(b-\tau)^4\Big]y(\tau)\mathrm{d}\tau\\
&+\int_t^b (t-\tau)\Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4\Big]y(\tau)\mathrm{d}\tau\\
&+\int_t^b \frac{(t-\tau)^2}{2}\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2\\
&\qquad\qquad + 4a_1(b-\tau)^3 + a_0(b-\tau)^4\Big]y(\tau)\mathrm{d}\tau\\
&-\int_t^b \frac{(t-\tau)^2}{2}(b-\tau)^4u(\tau)\mathrm{d}\tau
\end{aligned} \tag{2.72}$$

The closed form expression for $y^{(1)}(t)$ is obtained by adding equations (2.71) and (2.72) and dividing by $[(t-a)^4 + (b-t)^4]$.
For finding the expression for $y^{(2)}(t)$, equations (2.14) and (2.15) are integrated twice,giving:

$$\begin{aligned}
&(t-a)^{(4)}y^{(2)}(t) \\
&= \Big[8(t-a)^3 - a_3(t-a)^4\Big]y^{(1)}(t) \\
&+ \Big[36(t-a)^2 + 8a_3(t-a)^3 - a_2(t-a)^4\Big]y(t) \\
&+ \int_a^t \Big[96(\tau-a) - 36a_3(\tau-a)^2 + 8a_2(\tau-a)^3 - a_1(\tau-a)^4\Big]y(\tau)\mathrm{d}\tau \\
&+ \int_a^t (t-\tau)\Big[-24 + 24a_3(\tau-a) - 12a_2(\tau-a)^2 \\
&\qquad\qquad + 4a_1(\tau-a)^3 - a_0(\tau-a)^4\Big]y(\tau)\mathrm{d}\tau \\
&+ (t-\tau)(\tau-a)^4u(\tau)\mathrm{d}\tau
\end{aligned} \tag{2.73}$$

$$\begin{aligned}
&(b-t)^{(4)}y^{(2)}(t) \\
&= \Big[-8(b-t)^3 - a_3(b-t)^4\Big]y^{(1)}(t) \\
&+ \Big[-36(b-t)^2 - 8a_3(b-t)^3 - a_2(b-t)^4\Big]y(t) \\
&+ \int_t^b \Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4\Big]y(\tau)\mathrm{d}\tau \\
&+ \int_t^b (t-\tau)\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2 \\
&\qquad\qquad + 4a_1(b-\tau)^3 + a_0(b-\tau)^4\Big]y(\tau)\mathrm{d}\tau \\
&- (t-\tau)(b-\tau)^4u(\tau)\mathrm{d}\tau
\end{aligned} \tag{2.74}$$

The closed form expression for $y^{(2)}(t)$ is obtained by adding equations (2.73) and (2.74) and dividing by $[(t-a)^4 + (b-t)^4]$.

Integrating equations (2.14) and (2.15) once we get an expression for $y^{(3)}(t)$,

$$
\begin{aligned}
&(t-a)^{(4)}y^{(3)}(t)\\
&=\Big[4(t-a)^3-a_3(t-a)^4\Big]y^{(2)}(t)\\
&+\Big[-12(t-a)^2+4a_3(t-a)^3-a_2(t-a)^4\Big]y^{(1)}(t)\\
&+\Big[24(t-a)-12a_3(t-a)^2+4a_2(t-a)^3-a_1(t-a)^4\Big]y(t)\\
&+\int\limits_a^t\Big[-24+24a_3(\tau-a)-12a_2(\tau-a)^2+4a_1(\tau-a)^3-a_0(\tau-a)^4\Big]y(\tau)\mathrm{d}\tau\\
&+\int\limits_a^t(\tau-a)^4u(\tau)\mathrm{d}\tau
\end{aligned}
\tag{2.75}
$$

$$
\begin{aligned}
&(b-t)^{(4)}y^{(3)}(t)\\
&=\Big[-4(b-t)^3-a_3(b-t)^4\Big]y^{(2)}(t)\\
&+\Big[-12(b-t)^2-4a_3(b-t)^3-a_2(b-t)^4\Big]y^{(1)}(t)\\
&+\Big[-24(b-t)-12a_3(b-t)^2-4a_2(b-t)^3-a_1(b-t)^4\Big]y(t)\\
&+\int\limits_t^b\Big[24+24a_3(b-\tau)+12a_2(b-\tau)^2+4a_1(b-\tau)^3+a_0(b-\tau)^4\Big]y(\tau)\mathrm{d}\tau\\
&-\int\limits_t^b(b-\tau)^4u(\tau)\mathrm{d}\tau
\end{aligned}
\tag{2.76}
$$

The closed form expression for $y^{(3)}(t)$ is obtained by adding equations (2.75) and (2.76) and dividing $[(t-a)^4+(b-t)^4]$.

Chapter 3

Previous works on Kernel Based Parameter and State Estimation of LTI SISO Systems on Finite Interval

In this chapter we define two different approaches and previous works for Parameter estimation of an LTI system using Double Sided Kernel on a finite interval$[a, b]$.The following specific approaches are chosen to show research progress and improvement over time by overcoming limitations.

3.1 Two-step Non-asymptotic Parameter and State Estimation [18]

In this method a completely novel approach to two-step parameter estimation and state estimation from single noisy realizations of the system output on a finite interval $[a, b]$ is derived.

3.1.1 Step 1: Parameter estimation [18]

The parameter estimation problem to "identify" the true parameters $a0$,$a1$,$a2$,$a3$ in terms of simple unconstrained minimization is formulated using a simple approach of averaging the reproducing property itself(refer to Theorem 1 in chapter 2) . If exact matching is required only at a finite number n of discrete time points $t_j; j = 1, \cdots, n$ then the problem amounts to finding the optimal solution to:

$$\min\{J(a) := \frac{1}{2n}\sum_{i=1}^{n}(y(t_i)- < y, K_{DS}(t_i,\cdot) >_2)^2 \mid \text{w.r.t. } a \in \mathbb{R}^3\} \tag{3.1}$$
$$= \min\{\frac{1}{2n}\sum_{i=1}^{n}\left[y(t_i) - \int_a^b K_{DS}(t_i,\tau)y(\tau)d\tau\right]^2 \mid \text{w.r.t. } a \in \mathbb{R}^3\}$$

A continuous time version of the above becomes

$$\min\{\frac{1}{2T}\int_a^b \left[y(t) - \int_a^b K_{DS}(t,\tau)y(\tau)d\tau\right]^2 dt \mid \text{w.r.t. } a \in \mathbb{R}^3\} \tag{3.2}$$

with $T := b - a$.

The cost function in (3.2) can be calculated as follows. $K_{DS}(t,\tau)$ is expressed as a scalar product of some partial kernels

$$K_{DS}(t,\tau) = K_v(t,\tau)^T a + k_{v5}(t,\tau) \text{ equivalently } K_{DS}(t,\tau) = a^T K_v(t,\tau) + k_{v5}(t,\tau) \tag{3.3}$$

$$K_v(t,\tau)^T := [k_{v1}(t,\tau), k_{v2}(t,\tau), k_{v3}(t,\tau), k_{v5}(t,\tau)];\ a := [a_0, a_1, a_2, a_3]^T \tag{3.4}$$

Substituting the above into the cost of (3.2) yields (with $T := b - a$)

$$J(a) := \frac{1}{2T}\int_a^b \left[y(t) - \int_a^b K_{DS}(t,\tau)\ y(\tau)d\tau\right]^2 dt$$
$$= \frac{1}{T}\int_a^b \left[\frac{1}{2}y(t)^2 - y(t)\int_a^b K_{DS}(t,\tau)\ y(\tau)d\tau + \frac{1}{2}\left(\int_a^b K_{DS}(t,\tau)\ y(\tau)d\tau\right)^2\right] dt$$
$$= \frac{1}{T}\int_a^b \left[\frac{1}{2}y(t)^2 - y(t)\int_a^b K_{DS}(t,\tau)\ y(\tau)d\tau + \frac{1}{2}\int_a^b K_{DS}(t,\tau)\ y(\tau)d\tau\int_a^b K_{DS}(t,s)\ y(s)ds\right] dt$$
$$= \frac{1}{2T}\int_a^b y(t)^2 dt - \frac{1}{T}\int_a^b\int_a^b K_{DS}(t,\tau)\ y(\tau)y(t)\ d\tau\ dt$$
$$+ \frac{1}{2T}\int_a^b\int_a^b\int_a^b K_{DS}(t,\tau)K_{DS}(t,s)\ y(\tau)y(s)\ d\tau\ ds\ dt$$

Hence

$$
\begin{aligned}
&J(a) = \\
&= \frac{1}{2T}\int_a^b y(t)^2 dt - \frac{1}{T}\int_a^b\int_a^b [K_v(t,\tau)^T a + k_{v5}(t,\tau)]\ y(\tau)y(t)\ d\tau\ dt \\
&\quad + \frac{1}{2T}\int_a^b\int_a^b\int_a^b [a^T K_v(t,\tau) + k_{v5}(t,\tau)][K_v(t,s)^T a + k_{v5}(t,s)]\ y(\tau)y(s)\ d\tau\ ds\ dt
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2T}\int_a^b y(t)^2 dt - \frac{1}{T}\int_a^b\int_a^b K_v(t,\tau)^T\ y(\tau)y(t)\ d\tau\ dt\ a - \frac{1}{T}\int_a^b\int_a^b k_{v5}(t,\tau)\ y(\tau)y(t)\ d\tau\ dt \\
&+ \frac{1}{2T}\ a^T \int_a^b\int_a^b \left[\int_a^b K_v(t,\tau)K_v(t,s)^T\ dt\right] y(\tau)y(s)\ d\tau\ ds\ a \\
&+ \frac{1}{2T}\int_a^b\int_a^b\int_a^b k_{v5}(t,\tau)K_v(t,s)^T\ y(\tau)y(s)\ d\tau\ ds\ dt\ a \\
&+ a^T\frac{1}{2T}\int_a^b\int_a^b\int_a^b K_v(t,\tau)k_{v5}(t,s)\ y(\tau)y(s)\ d\tau\ ds\ dt \\
&+ \frac{1}{2T}\int_a^b\int_a^b\int_a^b k_{v5}(t,\tau)k_{v5}(t,s)]\ y(\tau)y(s)\ d\tau\ ds\ dt
\end{aligned}
$$

Assembling terms

$$
\begin{aligned}
&J(a) = d + b^T a + \frac{1}{2}a^T C a \quad \text{with} \\
&d := \left\{\frac{1}{2T}\int_a^b y(t)^2 dt - \frac{1}{T}\int_a^b\int_a^b k_{v5}(t,\tau)\ y(\tau)y(t)\ d\tau\ dt \right.\\
&\qquad\qquad \left. + \frac{1}{2T}\int_a^b\int_a^b\int_a^b k_{v5}(t,\tau)k_{v5}(t,s)]\ y(\tau)y(s)\ d\tau\ ds\ dt\right\} \\
&b^T := \left\{-\frac{1}{T}\int_a^b\int_a^b K_v(t,\tau)^T\ y(\tau)y(t)\ d\tau\ dt \right.\\
&\qquad\qquad \left. + \frac{1}{2T}\int_a^b\int_a^b\int_a^b [K_v(t,s)^T k_{v5}(t,\tau) + K_v(t,\tau)^T k_{v5}(t,s)]\ y(\tau)y(s)\ d\tau\ ds\ dt\right\} \\
&C := \frac{1}{T}\left\{\int_a^b\int_a^b \left[\int_a^b K_v(t,\tau)K_v(t,s)^T\ dt\right] y(\tau)y(s)\ d\tau\ ds\right\}
\end{aligned}
$$

The standard quadratic cost yields a minimization problem w.r.t. parameter a that is solved globally and analytically; [24] :

$$\begin{aligned} &J(a) := d + b^T a + \frac{1}{2} a^T C a \\ &\min\{J(a) \mid a \in \mathbb{R}^3\} \quad \text{is attained globally and uniquely at} \\ &\hat{a} = -C^{-1} b; \quad \text{with minimum value} \;\; J(\hat{a}) = d - \frac{1}{2} b^T C^{-1} b \end{aligned} \tag{3.5}$$

Also, it should be noted that the triple intergrals can be written as alternative integral products expressions which are easier to handle numerically

$$\int_a^b \int_a^b \int_a^b k_{v5}(t,\tau) k_{v5}(t,s)] \; y(\tau) y(s) \; d\tau \; ds \; dt = \int_a^b \left(\int_a^b k_{v5}(t,\tau) \; y(\tau) \; d\tau \right)^2 dt$$

$$\begin{aligned} &\int_a^b \int_a^b \int_a^b [K_v(t,s)^T k_{v5}(t,\tau) + K_v(t,\tau)^T k_{v5}(t,s)] \; y(\tau) y(s) \; d\tau \; ds \; dt \\ &\quad = \int_a^b \left(\int_a^b K_v(t,s)^T y(s) \; ds \right) \left(\int_a^b k_{v5}(t,\tau) \; y(\tau) \; d\tau \right) \; dt \\ &\qquad + \int_a^b \left(\int_a^b K_v(t,\tau)^T y(\tau) \; d\tau \right) \left(\int_a^b k_{v5}(t,s) \; y(s) \; ds \right) \; dt \end{aligned}$$

$$\begin{aligned} &\int_a^b \int_a^b \left[\int_a^b K_v(t,\tau) K_v(t,s)^T \; dt \right] y(\tau) y(s) \; d\tau \; ds \\ &\quad = \int_a^b \left[\int_a^b K_v(t,\tau) \; y(\tau) \; d\tau \right] \left[\int_a^b K_v(t,s) \; y(s) \; ds \right]^T \; dt \end{aligned}$$

The discrete cost (3.1) can be computed similarly, as follows:

$$
\begin{aligned}
J(a) &:= \frac{1}{2n}\sum_{i=1}^{n}\left[y(t_i) - \int_a^b K_{DS}(t_i,\tau)\ y(\tau)d\tau\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2}y(t_i)^2 - y(t_i)\int_a^b K_{DS}(t_i,\tau)\ y(\tau)d\tau + \frac{1}{2}\left(\int_a^b K_{DS}(t_i,\tau)\ y(\tau)d\tau\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2}y(t_i)^2 - y(t_i)\int_a^b K_{DS}(t_i,\tau)\ y(\tau)d\tau + \frac{1}{2}\int_a^b K_{DS}(t_i,\tau)\ y(\tau)d\tau\int_a^b K_{DS}(t_i,s)\ y(s)ds\right] \\
&= \frac{1}{2n}\sum_{i=1}^{n}y(t_i)^2 - \frac{1}{n}\sum_{i=1}^{n}y(t_i)\int_a^b K_{DS}(t_i,\tau)\ y(\tau)\ d\tau \\
&\qquad + \frac{1}{2n}\sum_{i=1}^{n}\int_a^b\int_a^b K_{DS}(t_i,\tau)K_{DS}(t_i,s)\ y(\tau)y(s)\ d\tau\ ds
\end{aligned}
$$

Expanding the kernels yields

$$
\begin{aligned}
&J(a) = \\
&= \frac{1}{2n}\sum_{i=1}^{n}y(t_i)^2 - \frac{1}{n}\sum_{i=1}^{n}\int_a^b [K_v(t_i,\tau)a + k_{v5}(t_i,\tau)]\ y(\tau)y(t_i)\ d\tau \\
&\qquad + \frac{1}{2n}\sum_{i=1}^{n}\int_a^b\int_a^b [a^T K_v(t_i,\tau) + k_{v5}(t_i,\tau)][K_v(t_i,s)^T a + k_{v5}(t_i,s)]\ y(\tau)y(s)\ d\tau\ ds \\
&= \frac{1}{2n}\sum_{i=1}^{n}y(t_i)^2 dt - \frac{1}{n}\sum_{i=1}^{n}\int_a^b K_v(t_i,\tau)^T\ y(\tau)y(t_i)\ d\tau\ a - \frac{1}{n}\sum_{i=1}^{n}\int_a^b k_{v5}(t_i,\tau)\ y(\tau)y(t_i)\ d\tau \\
&+ \frac{1}{2n}a^T\sum_{i=1}^{n}\int_a^b\int_a^b K_v(t_i,\tau)K_v(t_i,s)^T y(\tau)y(s)\ d\tau\ ds\ a \\
&+ \frac{1}{2n}\sum_{i=1}^{n}\int_a^b\int_a^b k_{v5}(t_i,\tau)K_v(t_i,s)^T\ y(\tau)y(s)\ d\tau\ ds\ a \\
&+ a^T\frac{1}{2n}\sum_{i=1}^{n}\int_a^b\int_a^b K_v(t_i,\tau)k_{v5}(t_i,s)\ y(\tau)y(s)\ d\tau\ ds \\
&+ \frac{1}{2n}\sum_{i=1}^{n}\int_a^b\int_a^b k_{v5}(t_i,\tau)k_{v5}(t_i,s)]\ y(\tau)y(s)\ d\tau\ ds
\end{aligned}
$$

Assembling terms again delivers a standard quadratic

$$J(a) = d + b^T a + \frac{1}{2} a^T C a \quad \text{with}$$
$$d := \left\{ \frac{1}{2n} \sum_{i=1}^{n} y(t_i)^2 - \frac{1}{n} \sum_{i=1}^{n} \int_a^b k_{v5}(t_i, \tau)\ y(\tau) y(t_i)\ d\tau \right.$$
$$\left. + \frac{1}{2n} \sum_{i=1}^{n} \int_a^b \int_a^b k_{v5}(t_i,\tau) k_{v5}(t_i, s)]\ y(\tau) y(s)\ d\tau\ ds \right\}$$

$$b^T := \left\{ -\frac{1}{n} \sum_{i=1}^{n} \int_a^b K_v(t_i,\tau)^T\ y(\tau) y(t_i)\ d\tau \right.$$
$$\left. + \frac{1}{2n} \sum_{i=1}^{n} \int_a^b \int_a^b [K_v(t_i,s)^T k_{v5}(t_i,\tau) + K_v(t_i,\tau)^T k_{v5}(t_i,s)]\ y(\tau) y(s)\ d\tau\ ds \right\}$$

$$C := \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_a^b \int_a^b K_v(t_i,\tau) K_v(t_i,s)^T\ y(\tau) y(s)\ d\tau\ ds \right\}$$

$$\min\{J(a) \mid a \in \mathbb{R}^3\} \quad \text{is attained globally and uniquely at}$$
$$\hat{a} = -C^{-1} b; \quad \text{with minimum value } \ J(\hat{a}) = d - \frac{1}{2} b^T C^{-1} b \tag{3.6}$$

The double integrals above can again be written in terms of single integrals

$$\int_a^b \int_a^b k_{v5}(t_i,\tau) k_{v5}(t_i,s)]\ y(\tau) y(s)\ d\tau\ ds = \left(\int_a^b k_{v5}(t_i,\tau)\ y(\tau)\ d\tau \right)^2$$
$$\int_a^b \int_a^b [K_v(t_i,s)^T k_{v5}(t_i,\tau) + K_v(t_i,\tau)^T k_{v5}(t_i,s)]\ y(\tau) y(s)\ d\tau\ ds$$
$$= \left(\int_a^b K_v(t_i,s)^T y(s)\ ds \right) \left(\int_a^b k_{v5}(t_i,\tau)\ y(\tau)\ d\tau \right)$$

$$+\left(\int_a^b K_v(t_i,\tau)^T y(\tau)\ d\tau\right)\left(\int_a^b k_{v5}(t_i,s)\ y(s)\ ds\right)$$

$$\int_a^b\int_a^b K_v(t_i,\tau)K_v(t_i,s)^T y(\tau)y(s)\ d\tau\ ds$$

$$=\left[\int_a^b K_v(t_i,\tau)\ y(\tau)\ d\tau\right]\left[\int_a^b K_v(t_i,s)\ y(s)\ ds\right]^T$$

3.1.2 Step 2: Output estimation by projection [18]

The system output after estimating the parameters can be reconstructed using

$$\hat{y}=\int_a^b K_{DS}(t_i,\tau)z(\tau)d\tau \tag{3.7}$$

where $z(\tau)$ is the measured signal.
However, as a more precise alternative to the above, the system output can be smoothed/reconstructed from a noisy measurement by direct orthogonal projection onto the subspace spanned by the fundamental solutions of the characteristic equation (2.4) of the system. This is because every solution of the characteristic equation with the already identified parameter vector a satisfies the reproducing property of Theorem 1 in chapter 2 and so, the projection onto the space of fundamental solutions will be the noise free trajectory of the system. The fundamental solutions of the characteristic equation are obtained by the direct integration of (2.4). This is performed as follows. We select n independent vectors as initial conditions for the homogeneous LTI system in (2.4) in chapter 2. These initial conditions can be taken as the vectors of the canonical basis in $\mathbb{R}^n$ i.e,

$$e_1=[1,0,...,0]$$
$$e_2=[0,1,...,0]$$
$$\cdots \tag{3.8}$$
$$e_n=[0,0,...,1] \tag{3.9}$$

The system equation (2.4) is then solved for each individual initial condition yielding solutions

$$\overline{y_i}(a)=e_i \qquad i=1,\cdots,n.$$

It is an elementary fact from the theory of ordinary differential equations that any solution of the system (2.4) with any initial condition is a linear combination of such fundamental solutions $\overline{y_i}$. Hence we search for the coefficients of this linear combination so that the resulting function is the closest to the output measurement data. Closest solution is found in terms of an orthogonal projection onto span of S^a.

$$S^a = \text{span}\left\{\overline{y}_i(\cdot), i = 1, \cdots, n\right\}$$

This is best done by orthonormalizing the set of fundamental solutions.The projection of a measured noisy signal $z(\cdot) \in L^2[a,b]$ into S^a is given as,

$$y_E(\cdot) \triangleq \arg\min\left\{\|z-y\|_2^2 \,|y \in S^a\right\} \tag{3.10}$$

We seek,

$$\hat{y} = \sum_{i=1}^{n} \hat{c}_i \overline{y}_i \tag{3.11}$$

As $\hat{y}$ is a linear combination, the optimality conditions in (3.10) is achieved if and only if,

$$\left\langle z|\overline{y}_j\right\rangle_2 = \sum_{i=1}^{n} \hat{c}_i \left\langle \overline{y}_i|\overline{y}_j\right\rangle_2 j = 1, \cdots, n \tag{3.12}$$

which can be written in a matrix form as:

$$\begin{aligned} v &= G(\overline{y})\hat{c}; \quad G(\overline{y}) \triangleq \text{mat}\left\{\left\langle \overline{y}_i|\overline{y}_j\right\rangle_2\right\}_{i,j=1}^{n} \\ v &\triangleq \text{vec}\left\{\left\langle z|\overline{y}_i\right\rangle_2\right\}_{i=1}^{n}; \hat{c} \triangleq \text{vec}\left\{\hat{c}_i\right\}_{i=1}^{n} \end{aligned} \tag{3.13}$$

G is called the Gram matrix for vectors in span S^a and is invertible because it is known that all fundamental solutions are linearly independent, from the theory of differential equations.

$$\hat{c} = G^{-1}(\overline{y})v \tag{3.14}$$

$\hat{y}$, the estimated output is thus obtained from (3.11).

3.1.3 Reconstruction of the output derivatives [18]

Once we obtain $\hat{y}$, the estimated output, the derivatives can be reconstructed using the formula [15],

$$y^{(i)}(t) = \int_a^b K_{DS}^i(t,\tau)\hat{y}(\tau)\mathrm{d}\tau \tag{3.15}$$

where, K_{DS}^i are the kernel representation for output derivatives. In this thesis we consider $i = 1, 2$, and 3. The formulae for kernel representation of output derivatives are developed in Chapter 2.

3.2 Parametric estimation and State Estimation using IV-GLS method [12]

3.2.1 Parametric estimation as a least squares problem [12]

Consider the input-output equation (2.5) of an n^{th} order

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) = -b_{n-1}u^{(n-1)}(t) - \cdots - b_0 u(t)$$

For a given input function u, the estimation problem for system reduces to that of a homogeneous system as the influence of the input can be factored out from the output measurement prior to the estimation procedure. The double-sided kernel $K_{DS,y}$ are clearly linear with respect to the system parameters, So as the kernels of Theorem 1 are linear in the unknown system coefficients, the reproducing property is first re-written to bring out this fact while omitting the obvious dependence of the kernels on n. [11].

$$y(t) = \int_a^b K_{DS,y}(t,\tau)y(\tau)\ d\tau \tag{3.16}$$

$$= \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i),y}(t,\tau)y(\tau)d\tau \tag{3.17}$$

where the $K_{DS(i),y}; i = 0, \cdots, n$ are "component kernels" of $K_{DS,y}$ that post-multiply the coefficients $\beta_i = a_i; i = 0, \cdots, n-1$, with $\beta_n = 1$ for convenience of notation. In a noise-free deterministic setting, the output variable y becomes the measured output coinciding with the nominal output trajectory y_T, so the regression equation for the constant parameters $a_i, i = 0, \ldots, n-1$, (3.17), can be written in a partitioned form as

$$y_T(t) = [K^{\bar{a}}, K^1](t; y_T)\beta^T \tag{3.18}$$

$$\bar{a} := [a_0, \cdots, a_{n-1}]; \quad \beta^T := [\bar{a}; \beta_n]$$

where $K^{\bar{a}}(t; y_T)$ is a row vector with integral components

$$K^{\bar{a}}(t; y_T)_k := \int_a^b K_{DS(k),y}(t,\tau)y_T(\tau)d\tau\ ;\ k = 0, \ldots n-1 \tag{3.19}$$

while $K^1(t; y_T)$ is a scalar

$$K^1(t; y_T) := \int_a^b K_{DS(n),y}(t,\tau) y_T(\tau) d\tau \tag{3.20}$$

corresponding to $\beta_n = 1$. Given distinct time instants $t_1, \cdots, t_N \in (a, b]$, here referred to as *knots*, the regression equation is re-written point-wise in the form of a matrix equation

$$Q(y_T) = P(y_T)\bar{a} \tag{3.21}$$

$$Q := \begin{bmatrix} q(t_1) \\ \vdots \\ q(t_N) \end{bmatrix} \quad ; P := \begin{bmatrix} p_0(t_1) \cdots p_{n-1}(t_1) \\ \ddots \\ p_0(t_N) \cdots p_{n-1}(t_N) \end{bmatrix}$$

$$q(t_i) = y_T(t_i) - K^1(t_i, y_T);$$

$$p_k(t_i) = K^{\bar{a}}(t_i; y_T)_k \tag{3.22}$$

that can be solved using least squares error minimization. provided adequate identifiability assumptions are met and the output is measured without error.Equation 3.21 is the parametric estimation equation to be solved by least squares error minimization.

3.2.2 Instrument Variable - Generalised Least Squares Method (IV - GLS) [12]

The formulation of the simple regression with distinct time instants $t_1, \cdots, t_N \in (a, b]$, referred to as *knots* is similar to that of equation (3.21). However least squares error minimisation can be solved if the identifiability assumptions are met.The identifiablity conditions are discussed elaborately in the next chapter.With the presence of large measurement noise, here assumed to be AWGN - white Gaussian and additive, the regression equation (3.21) is no longer valid as the reproducing property fails to hold along an inexact output trajectory. It must thus be suitably replaced leading to a stochastic regression problem. The stochastic output measurement process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is

$$y_M(t, \omega) = y_T(t) + \eta^\sigma(t, \omega) \ ; \quad t \in [a, b] \tag{3.23}$$

where η^σ signifies the AWGN with constant variance σ^2 and where y_T is the true system output. Here, W^σ is the Wiener process which is used to represent the integral of white noise , informally, $\eta^\sigma(t)dt = \sigma dW(t)$ with W as the standard Brownian motion. It follows that the following equality is valid

$$y_M(t) = \int_a^b K_{DS,y}(t,\tau) y_M(\tau)\ d\tau + e(t) \tag{3.24}$$

$$\text{with}\ \ e(t) := \eta^\sigma(t) - \int_a^b K_{DS,y}(t,\tau)\ dW^\sigma(\tau) \tag{3.25}$$

Therefore (3.17) becomes

$$y_M(t) = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i),y}(t,\tau) y_M(\tau) d\tau + e(t) \tag{3.26}$$

The above linear regression equation (3.26) fails to agree with the Gauss-Markov assumptions as the regressor is correlated with error which is also heteroskedastic.Hetero-skedastic error is an error with unequal variance. A standard quite powerful way to deal with *unknown* heteroskedasticity is to resort to generalized least squares (GLS), which can be shown to be BLUE ; see [1]. Let $Q(y_M)$ and $P(y_M)$ be the matrices corresponding to N samples of the measurement process realization y_M at a batch of knots $t_1, \cdots, t_N$. Then the stochastic regression error vector is given by

$$e := [e(t_i), \cdots, e(t_N)]^T = Q(y_M) - P(y_M)\bar{a} \tag{3.27}$$

where $e(t_i)$ are as in (3.25).
Applying the expectation operator to equations (3.23) and (3.25) and using the properties of the Wiener process yields

$$\mathrm{E}[y_M(t)] = \mathrm{E}[y_T(t)] + \mathrm{E}[\eta^\sigma(t)] = y_T(t) \tag{3.28}$$

$$\mathrm{E}[e(t)] = \mathrm{E}[\eta^\sigma(t)] - \mathrm{E}\left[\int_a^b K_{DS}(t,\tau) dW^\sigma(\tau)\right] \tag{3.29}$$

$$= \int_a^b K_{DS}(t,\tau)\mathrm{E}[dW^\sigma(\tau)] = 0 \tag{3.30}$$

thus

$$\mathrm{Cov}(e) = \mathrm{E}\left[ee^T\right] \tag{3.31}$$

The components of the covariance matrix are calculated as $\mathrm{Cov}(e(t_i), e(t_j))$.The covariance formula for the generalized white noise process is written as

$$\mathrm{E}[\eta^\sigma(t)\eta^\sigma(s)] = \sigma^2\delta(t-s) \tag{3.32}$$

The sifting property (also referred to as sampling property) of the delta Dirac function, which is valid for all tempered distributions f (thus also functions with compact support which are square integrable), is needed here as stated in the form

$$\int_a^b f(t)\delta(t-s)dt = f(s) \tag{3.33}$$

Writing $\eta^\sigma(\tau)d\tau$ in place of $dW^\sigma(\tau)$, the covariance calculation is as follows

$$\begin{aligned}
&\mathrm{Cov}[e(t_i), e(t_j)] = \mathrm{E}[e(t_i)e(t_j)] \\
&= \mathrm{E}\left[\left[\eta^\sigma(t_i) - \int_a^b K_{DS}(t_i,\tau)\eta^\sigma(\tau)d\tau\right]\left[\eta^\sigma(t_j) - \int_a^b K_{DS}(t_j,s)\eta^\sigma(s)ds\right]\right] \\
&= \mathrm{E}[\eta^\sigma(t_i)\eta^\sigma(t_j)] - \mathrm{E}\left[\int_a^b K_{DS}(t_i,\tau)\eta^\sigma(t_j)\eta^\sigma(\tau)d\tau\right] - \mathrm{E}\left[\int_a^b K_{DS}(t_j,s)\eta^\sigma(t_i)\eta^\sigma(s)ds\right] \\
&\qquad + \mathrm{E}\left[\int_a^b\int_a^b K_{DS}(t_i,\tau)K_{DS}(t_j,s)\eta^\sigma(\tau)\eta^\sigma(s)d\tau ds\right] \\
&= \sigma^2\delta(t_i - t_j) - \int_a^b K_{DS}(t_i,\tau)\mathrm{E}[\eta^\sigma(t_j)\eta^\sigma(\tau)]d\tau - \int_a^b K_{DS}(t_j,s)\mathrm{E}[\eta^\sigma(t_i)\eta^\sigma(s)]ds \\
&\qquad + \int_a^b\int_a^b K_{DS}(t_i,\tau)K_{DS}(t_j,s)\mathrm{E}[\eta^\sigma(\tau)\eta^\sigma(s)]d\tau ds \\
&= \sigma^2\delta(t_i - t_j) - \sigma^2\int_a^b K_{DS}(t_i,\tau)\delta(\tau - t_j)d\tau - \sigma^2\int_a^b K_{DS}(t_j,\tau)\delta(\tau - t_i)d\tau \\
&\qquad + \sigma^2\int_a^b K_{DS}(t_i,\tau)\int_a^b K_{DS}(t_j,s)\delta(s-\tau)ds\ d\tau \\
&= \sigma^2\delta(t_i - t_j) - \sigma^2 K_{DS}(t_i,t_j) - \sigma^2 K_{DS}(t_j,t_i) + \sigma^2\int_a^b K_{DS}(t_i,\tau)K_{DS}(t_j,\tau)d\tau
\end{aligned}$$

The formula for the variance follows by substituting $t = t_i = t_j$, and recalling that, informally, $\delta(0) = 1$,

$$\mathrm{Var}[e(t)] = \sigma^2 - 2\sigma^2 K_{DS}(t,t) + \sigma^2 \int_a^b \big[K_{DS}(t,\tau)\big]^2 d\tau$$

At this point it should be clear that the standard GLS cannot be applied directly as the covariance matrix depends on the unknown variance σ^2 and also on the unknown parameter vector $\bar{a}$ in the K_{DS} kernels. Hence a feasible version of the GLS must be employed here in which the covariance matrix is estimated progressively as more information about the regression residuals becomes available. This is typically performed as part of a recursive scheme in which consecutive batches of samples are drawn from the realization of y_M. Letting $Q_i - P_i\bar{a}$ denote the regression error e_i in batch i, the recursive GLS algorithm computes

$$\hat{a}_k = \arg\min_{\bar{a}} \left(\sum_{i=1}^{k} (Q_i - P_i\bar{a})^T S_i (Q_i - P_i\bar{a}) \right) \tag{3.34}$$

where $\hat{a}_k$ is the parameter estimate update at iteration k of the algorithm. Each weighting matrix S_{k+1}, is calculated as the inverse of the covariance matrix based on the parameter estimate $\hat{a}_k$ obtained.The Recursive GLS with covariance weighting with detailed steps is described in next chapter(refer to section 4.4)

3.2.3 Errors-in-variables [12]

It is well known that the presence of errors-in-variables induces an asymptotic bias in OLS regression estimates which is proportional to the signal-to-noise ratio in the observed regressand. In such situations the leading way to eliminate estimation bias is to use *Instrumental Variables* (IV); see [30], in the normal equations that deliver the optimal estimates. The IV method has a long history and multiple applications; refer to [23], [28], [29], [3], [10], [4].

To render statistical consistency for the estimation problem at hand the IV is constructed by way of the backward reproducing kernel as described below.

Referring to the exposition of the basic regression problem in section 3.2.1 it follows from Theorem 1 that the “double-sided” regression equation (3.18), can be cloned as two statistically independent regression equations corresponding to the forward and backward kernels $K_{F,y}$ and $K_{B,y}$ as follows:

$$\begin{aligned}(t-a)^n y_T(t) &= \int_a^t K_{F,y}(t,\tau)\; y_T(\tau)d\tau \\ &:= [K_F^{\bar{a}}, K_F^1](t; y_T)\;\beta\;; t\in[a,b] \qquad (3.35)\\ (b-t)^n y_T(t) &= \int_t^b K_{B,y}(t,\tau)\; y_T(\tau)d\tau \\ &:= [K_B^{\bar{a}}, K_B^1](t; y_T)\;\beta\;; t\in[a,b] \qquad (3.36)\end{aligned}$$

Given a set of knots $[t_1, \cdots, t_N]; N > n; t_1 >> a; t_N << b$, the latter are written in discrete time as N copies of (3.35) and (3.36) in matrix-vector form

$$\begin{aligned} Y_T &= K_F(y_T)\beta \qquad (3.37)\\ Y_T &= K_B(y_T)\beta \qquad (3.38)\\ Y_T &:= \begin{bmatrix} y_T(t_1) & \cdots & y_T(t_N)\end{bmatrix}^T \\ K_F(y_T) &:= \begin{bmatrix} \frac{1}{(t_1-a)^n}[K_F^{\bar{a}}, K_F^1](t_1, y_T) \\ \vdots \\ \frac{1}{(t_N-a)^n}[K_F^{\bar{a}}, K_F^1](t_N, y_T)\end{bmatrix}; \\ K_B(y_T) &:= \begin{bmatrix} \frac{1}{(b-t_1)^n}[K_B^{\bar{a}}, K_B^1](t_1, y_T) \\ \vdots \\ \frac{1}{(b-t_N)^n}[K_B^{\bar{a}}, K_B^1](t_N, y_T)\end{bmatrix}\end{aligned}$$

The equations (3.37) and (3.38) deliver two "independent" OLS estimators for the parameter vector β corresponding to two normal equations :

$$\begin{aligned} K_F(y_T)^T K_F(y_T)\beta &= K_F(y_T)^T Y_T \\ K_B(y_T)^T K_B(y_T)\beta &= K_B(y_T)^T Y_T\end{aligned}$$

namely:

$$\begin{aligned}\hat{\beta}_F &:= \left[K_F(y_T)^T K_F(y_T)\right]^{-1} K_F(y_T)^T Y_T \qquad (3.39)\\ \hat{\beta}_B &:= \left[K_B(y_T)^T K_B(y_T)\right]^{-1} K_B(y_T)^T Y_T \qquad (3.40)\end{aligned}$$

provided that each of the inverted matrices are nonsingular. The more efficient of them would be the one that corresponds to an inverted matrix having smaller condition number.

Of course, nothing stands in the way of pre-multiplying the forward estimation equation (3.37) by the backward matrix $K_B(y_T)$, or, vice-versa, pre-multiplying the backward estimation equation (3.38) by the forward matrix $K_F(y_T)$. If any of the symmetric matrices satisfy :

$$\det\left[K_B(y_T)^T K_F(y_T)\right] \neq 0 \quad \text{or}$$
$$\det\left[K_F(y_T)^T K_B(y_T)\right] \neq 0 \tag{3.41}$$

the other one will also be nonsingular (because they are transposes of one another). This observation delivers two more estimators which, in the noiseless case, will be competitive with those in (3.39)-(3.40) :

$$K_B(y_T)^T K_F(y_T)\beta = K_B(y_T)^T Y_T$$
$$K_F(y_T)^T K_B(y_T)\beta = K_F(y_T)^T Y_T$$

so that

$$\hat{\beta}_{IVF} := \left[K_B(y_T)^T K_F(y_T)\right]^{-1} K_B(y_T)^T Y_T \tag{3.42}$$
$$\hat{\beta}_{IVB} := \left[K_F(y_T)^T K_B(y_T)\right]^{-1} K_F(y_T)^T Y_T \tag{3.43}$$

These are in fact "instrumental variable estimators" as compared with the OLS estimators (3.39) and (3.40) with K_B as the IV for the forward equation (3.42) and with K_F as the IV for the backward equation (3.43). Clearly, as long as the non singularity condition (3.41) is satisfied then, in the noiseless case, all the estimators are bound to produce the same value of the estimated parameter vector, i.e. $\beta_F = \beta_B = \beta_{IVF} = \beta_{IVB}$.

The use of the IV as defined above is rigorously justified as follows. When a noisy realization of a measurement process y_M replaces the unknown system output y_T, the regression equations (3.37), (3.38) re-write as

$$Y_M = K_F(y_M)\beta + \xi \tag{3.44}$$

$$Y_M = K_B(y_M)\beta + \psi \tag{3.45}$$

with the obvious meaning of $K_F(y_M)$ and $K_F(y_M)$ whose arguments were substituted accordingly. As follows from the definition of the stochastic regression error (3.25) the noise processes ξ and ψ are expressed by

$$\xi := \eta - K_F(\eta)\beta; \quad \psi := \eta - K_B(\eta)\beta \tag{3.46}$$

Now, the "instrumental variable" $K_B(y_M)$ is well correlated with Y_M (as in the noise free case $K_B(y_T)$ produces a system output y_T that is computed by the forward kernel as well as by the backward kernel). On the other hand, the backward kernel matrix $K_B(y_M)$ is uncorrelated with the "forward" noise ξ (over the interval (a,t)) if the backward kernel $K_B(y_M)$ (involving the noise ψ) is calculated over a disjoint interval $(t+\varepsilon, b)$. Implied is the fact $\mathrm{E}[K_B(y_M)\psi] = 0$ while $\mathrm{E}[K_B(y_M)K_F(y_M)] \neq 0$ for a small separation constant $\varepsilon > 0$.

A symmetric statement is that the "instrumental variable" $K_F(y_M)$ is uncorrelated with the "backward" noise ψ.

The statistical properties of the IV instrument as discussed above restore consistency of the modified GLS estimator employing such IV instrument; see [30] for a proof in the OLS case. However, the IV-GLS recursions are now modified to

$$\hat{\beta}_{k+1} = \hat{\beta}_k + R_{k+1}\tilde{P}_{k+1}^T S_{k+1}(\tilde{Q}_{k+1} - \tilde{P}_{k+1}\hat{\beta}_k) \tag{3.47}$$

$$R_{k+1} = R_k - R_k\tilde{P}_{k+1}^T(S_{k+1}^{-1} + \tilde{P}_{k+1}R_k\tilde{P}_{k+1}^T)^{-1}\tilde{P}_{k+1}R_k \tag{3.48}$$

where the matrices P and Q are replaced by $\tilde{P}$ and $\tilde{Q}$

$$\tilde{P} = K_B(y_M)^T K_F(y_M)\ ;\ \tilde{Q} = K_B(y_M)^T Y_M \tag{3.49}$$

and $M_{k+1} = \sum_{i=0}^{k+1} \tilde{P}_i^T S_i \tilde{P}_i$ and $R_{k+1} = M_{k+1}^{-1}$.

It is worth noticing that a stopping criterion of the recursive scheme is elegantly delivered by the fact that the final estimate should satisfy $\beta_n = 1$ where $\beta := [a_0, \cdots, a_{n-1}, \beta_n]$, lending a criterion

$$|\hat{a}_n - 1| < \epsilon \text{ for some } \epsilon > 0 \tag{3.50}$$

A similar recursive scheme for the estimator for $\hat{\beta}_{IVB}$ can also be developed.
When using the recursive IV-GLS algorithm, the covariance of the error terms should be calculated by replacing K_{DS} by K_F or K_B appropriately.

Chapter 4

Kernel-Based FGLS for Parameter Estimation in LTI Systems

The Parameter Estimation methods which uses Generalised Least Squares-IV method mentioned in Chapter 3 uses a simple regression equation (3.21).In our method we use Feasible Generalised Least Squares with multiple regression. Using Theorem 1, the kernel is integrated multiple times forming multiple linearly independent regression equations.Consider a fourth order system (2.13) which has four unknown parameters $a0, a1, a2, a3$ .The kernel is integrated four times forming four linearly independent stochastic regression equations matching the number of unknown parameters.These independent regression equations are derived below:

Consider the forward kernel representation (2.44) for simplicity,

$$\alpha_a(t)y(t) \triangleq \int_a^t K_F(t,\tau)y(\tau)d\tau \tag{4.1}$$

and using the Cauchy formula for repeated integration on (4.1) yields

$$\frac{1}{(n-1)!}\int_a^t \alpha_a(\tau)(t-\tau)^{n-1}y(\tau)d\tau = \frac{1}{n!}\int_a^t (t-\tau)^n K_F(t,\tau)y(\tau)d\tau \quad n=1,2,3 \tag{4.2}$$

from (2.45) we have $K_F(t,\tau)$ as,

$$
\begin{aligned}
K_F(t,\tau) = & \Big[16(\tau-a)^3 - a_3(\tau-a)^4\Big] \\
& + (t-\tau)\Big[-72(\tau-a)^2 + 12a_3(\tau-a)^3 - a_2(\tau-a)^4\Big] \\
& + \frac{(t-\tau)^2}{2}\Big[96(\tau-a) - 36a_3(\tau-a)^2 + 8a_2(\tau-a)^3 - a_1(\tau-a)^4\Big] \\
& + \frac{(t-\tau)^3}{6}\Big[-24 + 24a_3(\tau-a) - 12a_2(\tau-a)^2 + 4a_1(\tau-a)^3 - a_0(\tau-a)^4\Big]
\end{aligned}
\tag{4.3}
$$

substituting (4.3) in the (4.2) we get,

$$
\begin{aligned}
= & \int_a^t \frac{(t-\tau)^n}{n!}\Big[16(\tau-a)^3 - a_3(\tau-a)^4\Big]y(\tau)d\tau \\
+ & \int_a^t \frac{(t-\tau)^{n+1}}{(n+1)!}\Big[-72(\tau-a)^2 + 12a_3(\tau-a)^3 - a_2(\tau-a)^4\Big]y(\tau)d\tau \\
+ & \int_a^t \frac{(t-\tau)^{n+2}}{(n+2)!}\Big[96(\tau-a) - 36a_3(\tau-a)^2 + 8a_2(\tau-a)^3 - a_1(\tau-a)^4\Big]y(\tau)d\tau \\
+ & \int_a^t \frac{(t-\tau)^{n+3}}{(n+3)!}\Big[-24 + 24a_3(\tau-a) - 12a_2(\tau-a)^2 + 4a_1(\tau-a)^3 - a_0(\tau-a)^4\Big]y(\tau)d\tau
\end{aligned}
\tag{4.4}
$$

or,

$$
\frac{1}{(n-1)!}\int_a^t \alpha_a(\tau)(t-\tau)^{n-1}y(\tau)d\tau = \int_a^t K_F^n(t,\tau)y(\tau)d\tau \quad n = 1,2,3 \tag{4.5}
$$

where,

$$
\begin{aligned}
K_F^n(t,\tau) = & \frac{(t-\tau)^n}{n!}\Big[16(\tau-a)^3 - a_3(\tau-a)^4\Big] \\
& + \frac{(t-\tau)^{n+1}}{(n+1)!}\Big[-72(\tau-a)^2 + 12a_3(\tau-a)^3 - a_2(\tau-a)^4\Big] \\
& + \frac{(t-\tau)^{n+2}}{(n+2)!}\Big[96(\tau-a) - 36a_3(\tau-a)^2 + 8a_2(\tau-a)^3 - a_1(\tau-a)^4\Big] \\
& + \frac{(t-\tau)^{n+3}}{(n+3)!}\Big[-24 + 24a_3(\tau-a) - 12a_2(\tau-a)^2 + 4a_1(\tau-a)^3 - a_0(\tau-a)^4\Big]
\end{aligned}
\tag{4.6}
$$

A similar procedure is then used for the backward kernel representation.

$$\alpha_b(t)y(t) = \int_t^b K_B(t,\tau)y(\tau)d\tau = -\int_b^t K_B(t,\tau)y(\tau)d\tau \tag{4.7}$$

using the Cauchy formula for repeated integration on (4.7) yields

$$\frac{1}{(n-1)!}\int_t^b \alpha_b(\tau)(t-\tau)^{n-1}y(\tau)d\tau = -\frac{1}{n!}\int_t^b (t-\tau)^n K_B(t,\tau)y(\tau)d\tau \quad n = 1,2,3 \tag{4.8}$$

from (2.68) we have

$$\begin{aligned} K_{B,y}(t,\tau) = &\Big[16(b-\tau)^3 + a_3(b-\tau)^4\Big] \\ &+ (t-\tau)\Big[72(b-\tau)^2 + 12a_3(b-\tau)^3 + a_2(b-\tau)^4\Big] \\ &+ \frac{(t-\tau)^2}{2}\Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4\Big] \\ &+ \frac{(t-\tau)^3}{6}\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2 + 4a_1(b-\tau)^3 + a_0(b-\tau)^4\Big] \end{aligned} \tag{4.9}$$

substituting (4.9) in (4.8)

$$\begin{aligned} = -\Bigg[&\int_t^b \frac{(t-\tau)^n}{n!}\Big[16(b-\tau)^3 + a_3(b-\tau)^4\Big]y(\tau)d\tau \\ &+ \int_t^b \frac{(t-\tau)^{n+1}}{n+1!}\Big[72(b-\tau)^2 + 12a_3(b-\tau)^3 + a_2(b-\tau)^4\Big]y(\tau)d\tau \\ &+ \int_t^b \frac{(t-\tau)^{n+2}}{n+2!}\Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4\Big]y(\tau)d\tau \\ &+ \int_t^b \frac{(t-\tau)^{n+3}}{n+3!}\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2 + 4a_1(b-\tau)^3 + a_0(b-\tau)^4\Big]y(\tau)d\tau\Bigg] \end{aligned} \tag{4.10}$$

where,

$$
\begin{aligned}
K_B^n(t,\tau) = -\Bigg[&\frac{(t-\tau)^n}{n!}\Big[16(b-\tau)^3 + a_3(b-\tau)^4\Big] \\
&+ \frac{(t-\tau)^{n+1}}{n+1!}\Big[72(b-\tau)^2 + 12a_3(b-\tau)^3 + a_2(b-\tau)^4\Big] \\
&+ \frac{(t-\tau)^{n+2}}{n+2!}\Big[96(b-\tau) + 36a_3(b-\tau)^2 + 8a_2(b-\tau)^3 + a_1(b-\tau)^4\Big] \\
&+ \frac{(t-\tau)^{n+3}}{n+3!}\Big[24 + 24a_3(b-\tau) + 12a_2(b-\tau)^2 + 4a_1(b-\tau)^3 + a_0(b-\tau)^4\Big]\Bigg]
\end{aligned} \tag{4.11}
$$

or,

$$
\frac{1}{(n-1)!}\int_t^b \alpha_b(\tau)(t-\tau)^{n-1}y(\tau)d\tau = \int_t^b K_B^n(t,\tau)y(\tau)d\tau \quad n = 1,2,3 \tag{4.12}
$$

Therefore, adding equations (2.44) and (2.67)

$$
\alpha_a(t)y(t) + \alpha_b(t)y(t) = \int_a^t K_F(t,\tau)y(\tau)d\tau + \int_t^b K_B(t,\tau)y(\tau)d\tau \tag{4.13}
$$

$$
\alpha_a(t)y(t) + \alpha_b(t)y(t) = \int_a^b K_{DS}(t,\tau)y(\tau)d\tau \tag{4.14}
$$

with

$$
K_{DS}(t,\tau) \triangleq \begin{cases} K_F(t,\tau) : \tau \leq t \\ K_B(t,\tau) : \tau > t \end{cases} \tag{4.15}
$$

Adding (4.5) and (4.12) gives

$$
\begin{aligned}
\frac{1}{(n-1)!}\Bigg[\int_a^t \alpha_a(\tau)(t-\tau)^{n-1}y(\tau)d\tau + \int_t^b \alpha_b(\tau)(t-\tau)^{n-1}y(\tau)d\tau\Bigg] \\
= \Bigg[\int_a^t K'_F(t,\tau)y(\tau)d\tau + \int_t^b K'_B(t,\tau)y(\tau)\Bigg]
\end{aligned} \tag{4.16}
$$

$$\frac{1}{(n-1)!}\left[\int_a^t \alpha_a(\tau)(t-\tau)^{n-1}y(\tau)d\tau + \int_t^b \alpha_b(\tau)(t-\tau)^{n-1}y(\tau)d\tau\right] = \int_a^b K_{DS}^n(t,\tau)y(\tau)d\tau \quad n=1,2,3 \tag{4.17}$$

with

$$K_{DS}^n(t,\tau) \triangleq \begin{cases} K_F^n(t,\tau) : \tau \le t \\ K_B^n(t,\tau) : \tau > t \end{cases} \tag{4.18}$$

Substituting $n = 1, 2, 3$ in (4.17) and (4.13) yields four linearly independent equations with four unknown parameters.Using the kernel-based multiple regression equations parameter and states of a fourth order system are determined.

4.1 Parameter Estimation as a Least Squares Problem [12]

The kernels of Theorem 1 are linear in the unknown system coefficients β_i. Omitting the dependence of kernels on n the reproducing property (for homogeneous systems) is first re-written.The additional linearly independent equations formed by repeated integration as in 4.17 with $n = 1, 2, 3$ can also be represented as:

$$\alpha_{ab}(t)y(t) = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i)}(t,\tau)y(\tau)d\tau \tag{4.19}$$

$$\int_a^b \alpha_{ab}(\tau)y(\tau)d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i)}^1(t,\tau)y(\tau)d\tau \tag{4.20}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)y(\tau)d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i)}^2(t,\tau)y(\tau)d\tau \tag{4.21}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)^2y(\tau)d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i)}^3(t,\tau)y(\tau)d\tau \tag{4.22}$$

where the $K_{DS(i)}$ and $K_{DS(i)}^n$, $i = 0, ..., n$ are 'component kernels' of K_{DS} and K_{DS}^n respectively that post-multiply the coefficients $\beta_k = a_k$, $k = 0, ..., n-1$, with $\beta_n = 1$ for convenience of notation. In a noise-free deterministic setting, the output variable y becomes the measured output coinciding with the nominal output trajectory y_T, so the regression equations for the constant parameters a_k,$k = 0, ..., n-1$, can be written in a partitioned form as

$$\alpha_{ab}(t)y_T(t) = [K_{DS(k)}, K_{DS(n)}](t; y_T)\beta^T \tag{4.23}$$

$$\int_a^b \alpha_{ab}(\tau)y(\tau)d\tau = [K^1_{DS(k)}, K^1_{DS(n)}](t; y_T)\beta^T \tag{4.24}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)y(\tau)d\tau = [K^2_{DS(k)}, K^2_{DS(n)}](t; y_T)\beta^T \tag{4.25}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)^2 y(\tau)d\tau = [K_{DS(k)})^3, K^3_{DS(n)}](t; y_T)\beta^T \tag{4.26}$$

where

$$\overline{a} := [a_0, a_1, ..., a_{n-1}]; \beta^T := [\overline{a}, \beta_n]; k = 0, 1....n-1$$

where $K_{DS(k)}(t; y_T)$,$K^1_{DS(k)}(t; y_T)$,$K^2_{DS(k)}(t; y_T)$ and $K^3_{DS(k)}(t; y_T)$ are row vectors with integral components corresponding to $\overline{a}$ and $K_{DS(n)}$,$K^1_{DS(n)}$,$K^2_{DS(n)}$ and $K^3_{DS(n)}$ are scalars corresponding to $\beta_n = 1$.

$$\alpha_{ab}(t)y_T(t) = \overline{a}\int_a^b K_{DS(k)}(t,\tau)y(\tau)d\tau + \beta_n \int_a^b K_{DS(n)}(t,\tau)y(\tau)d\tau \tag{4.27}$$

$$\int_a^b \alpha_{ab}(\tau)y(\tau)d\tau = \overline{a}\int_a^b K^1_{DS(k)}(t,\tau)y(\tau)d\tau + \beta_n \int_a^b K^1_{DS(n)}(t,\tau)y(\tau)d\tau \tag{4.28}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)y(\tau)d\tau = \overline{a}\int_a^b K^2_{DS(k)}(t,\tau)y(\tau)d\tau + \beta_n \int_a^b K^2_{DS(n)}(t,\tau)y(\tau)d\tau \tag{4.29}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)^2 y(\tau)d\tau = \overline{a}\int_a^b K^3_{DS(k)}(t,\tau)y(\tau)d\tau + \beta_n \int_a^b K^3_{DS(n)}(t,\tau)y(\tau)d\tau \tag{4.30}$$

where

$$\overline{a} := [a_0, a_1, ..., a_{n-1}]; \beta_n = 1; k = 0, 1....n-1$$

taking the scalar components corresponding to β_n to the L.H.S

$$\alpha_{ab}(t)y_T(t) - \beta_n \int_a^b K_{DS(n)}(t,\tau)y(\tau)d\tau = \overline{a}\int_a^b K_{DS(k)}(t,\tau)y(\tau)d\tau \tag{4.31}$$

$$\int_a^b \alpha_{ab}(\tau)y(\tau)d\tau - \beta_n \int_a^b K^1_{DS(n)}(t,\tau)y(\tau)d\tau = \overline{a}\int_a^b K^1_{DS(k)}(t,\tau)y(\tau)d\tau \tag{4.32}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)y(\tau)d\tau - \beta_n \int_a^b K^2_{DS(n)}(t,\tau)y(\tau)d\tau = \overline{a}\int_a^b K^2_{DS(k)}(t,\tau)y(\tau)d\tau \tag{4.33}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)^2y(\tau)d\tau - \beta_n \int_a^b K^3_{DS(n)}(t,\tau)y(\tau)d\tau = \overline{a}\int_a^b K^3_{DS(k)}(t,\tau)y(\tau)d\tau \tag{4.34}$$

where

$$\overline{a} := [a_0, a_1, ..., a_{n-1}]; \beta_n = 1; k = 0, 1....n-1$$

the equations(4.31),(4.32), (4.33) and (4.34) can also be written in the matrix form .

$$Q(y_T) = P(y_T)\overline{a} \tag{4.35}$$

$$\begin{bmatrix} \alpha_{ab}(t)y_T(t) - \beta_n \int_a^b K_{DS(n)}(t,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)y(\tau)d\tau - \beta_n \int_a^b K^1_{DS(n)}(t,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t-\tau)y(\tau)d\tau - \beta_n \int_a^b K^2_{DS(n)}(t,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t-\tau)^2y(\tau)d\tau - \beta_n \int_a^b K^3_{DS(n)}(t,\tau)y(\tau)d\tau \end{bmatrix} = \begin{bmatrix} \overline{a}\int_a^b K_{DS(k)}(t,\tau)y(\tau)d\tau \\ \overline{a}\int_a^b K^1_{DS(k)}(t,\tau)y(\tau)d\tau \\ \overline{a}\int_a^b K^2_{DS(k)}(t,\tau)y(\tau)d\tau \\ \overline{a}\int_a^b K^3_{DS(k)}(t,\tau)y(\tau)d\tau \end{bmatrix}$$

that can be solved using least squares error minimization provided adequate identifiability assumptions are met and the output is measured without error.

4.1.1 Parameter estimation and Practical Linear Identifiablity

Given distinct time instants $t_i = t_1, \cdots, t_N \in (a, b]$, here referred to as *knots*, the regression equations are re-written point-wise in the form of a matrix equation

$$Q(y_T) = P(y_T)\overline{a} \tag{4.36}$$

$$\begin{bmatrix} \alpha_{ab}(t)y_T(t_i) - \beta_n \int_a^b K_{DS(n)}(t,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)y(\tau)d\tau - \beta_n \int_a^b K^1_{DS(n)}(t_i,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t_i-\tau)y(\tau)d\tau - \beta_n \int_a^b K^2_{DS(n)}(t_i,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t_i-\tau)^2 y(\tau)d\tau - \beta_n \int_a^b K^3_{DS(n)}(t_i,\tau)y(\tau)ds \end{bmatrix} = \begin{bmatrix} \overline{a}\int_a^b K_{DS(k)}(t_i,\tau)y(\tau)d\tau \\ \overline{a}\int_a^b K^1_{DS(k)}(t_i,\tau)y(\tau)d\tau \\ \overline{a}\int_a^b K^2_{DS(k)}(t_i,\tau)y(\tau)d\tau \\ \overline{a}\int_a^b K^3_{DS(k)}(t_i,\tau)y(\tau)d\tau \end{bmatrix}$$

where,

$$Q =: \begin{bmatrix} q^1(t_i) \\ q^2(t_i) \\ q^3(t_i) \\ q^4(t_i) \end{bmatrix} = \begin{bmatrix} \alpha_{ab}(t)y_T(t_i) - \beta_n \int_a^b K_{DS(n)}(t,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)y(\tau)d\tau - \beta_n \int_a^b K^1_{DS(n)}(t_i,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t_i-\tau)y(\tau)d\tau - \beta_n \int_a^b K^2_{DS(n)}(t_i,\tau)y(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t_i-\tau)^2 y(\tau)d\tau - \beta_n \int_a^b K^3_{DS(n)}(t_i,\tau)y(\tau)d\tau \end{bmatrix} \tag{4.37}$$

and

$$P =: \begin{bmatrix} p^1(t_i) \\ p^2(t_i) \\ p^3(t_i) \\ p^4(t_i) \end{bmatrix} = \begin{bmatrix} \int_a^b K_{DS(k)}(t_i,\tau)y(\tau)d\tau \\ \int_a^b K^1_{DS(k)}(t_i,s)y(\tau)d\tau \\ \int_a^b K^2_{DS(k)}(t_i,s)y(\tau)d\tau \\ \int_a^b K^3_{DS(k)}(t_i,\tau)y(\tau)d\tau \end{bmatrix}; \qquad \overline{a} = [a1, a2, a3, a4] \tag{4.38}$$

where,

$$q^1 := \begin{bmatrix} q^1(t_1) \\ \vdots \\ q^1(t_N) \end{bmatrix} \qquad p^1 := \begin{bmatrix} p^1_0(t_1)\cdots p^1_{n-1}(t_1) \\ \ddots \\ p^1_0(t_N)\cdots p^1_{n-1}(t_N) \end{bmatrix} \tag{4.39}$$

Identifiability of homogeneous LTI systems from a single realization of a measured output [12]

Referring to Definition 2.1 and Theorem 2.2 from [27], a homogeneous LTI system such as

$$\begin{aligned} &\dot{x}(t) = Ax(t); \quad y = Cx; \quad x \in \mathbb{R}^n \\ &x(0) = b \end{aligned} \tag{4.40}$$

is identifiable from a single noise-free realization of its output trajectory y under precise conditions, which admittedly are difficult to verify computationally. Stated in equivalent form:

Definition: [12] Model (4.40) *is globally identifiable from* b *if and only if the functional mapping* $b \mapsto y(\cdot; A, b)$ *is injective on* $\mathbb{R}^n$ *where* $y(\cdot; A, b)$ *denotes the output orbit of* (4.40) .

Theorem: [12] Model (4.40) *is globally identifiable from* b *if and only if the output orbit of* (4.40) *is not confined to a proper subspace of* $\mathbb{R}^n$.

An alternative theoretical identifiability criterion from a single *noisy* output trajectory that is easier to verify computationally will be presented elsewhere. The latter is stated in terms of linear independence of the functions involving the component kernel expressions in (3.17), namely functions:

$$f_i(t) := \int_a^b K_{DS(i),y}(t,\tau) y(\tau) d\tau \ ; \ i = 1, \cdots, n \tag{4.41}$$

Linear independence of the above set can be readily checked by establishing non-singularity of the Wronskian matrix for (4.41) at some point in the interval $[a, b]$. However, yet another *practical version of identifiability* is sufficient for the present estimation purpose as defined below.

Definition 2: Practical linear identifiability [12]

The homogeneous system (4.40) is practically linearly identifiable on $[a, b]$ with respect to a particular noisy discrete realization of the output measurement process, $y(t), t \in [a, b]$, if and only if there exist distinct knots $t_1, \cdots, t_N \in (a, b]$ which render $rank P(y) = n$. Any such output realization is then called *persistent*.

In the presence of large measurement noise, here assumed to be AWGN - white Gaussian and additive, the regression equation (3.17) is no longer valid as the reproducing property fails to hold along an inexact output trajectory. It must thus be suitably replaced leading to a stochastic regression problem. First, the stochastic output measurement process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is

$$y_M(t, \omega) = y_T(t) + \sigma \dot{W}(t) \; ; \quad t \in [a, b] \tag{4.42}$$

where $\sigma \dot{W}(t)$ signifies the AWGN with constant variance σ^2 and where y_T is the true system output. This corresponds to a random kernel expression

$$\int_a^b K_{DS,y}(t,\tau) y_M(\tau) \; d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i),y}(t,\tau) y_M(\tau) d\tau \tag{4.43}$$

which is better written using the proper stochastic nomenclature as

$$\int_a^b K_{DS,y}(t,\tau) y_T(\tau) \; d\tau + \int_a^b K_{DS,y}(t,\tau) \; dW^\sigma(\tau) \tag{4.44}$$

with its partitioned form following similarly. Here, W^σ is the Wiener process with intensity σ so that, informally, $\eta^\sigma(t)dt = \sigma dW(t)$ with W as the standard Brownian motion.since y_T satisfies the reproducing property in the deterministic regression equation (3.17). It is noted that the random error variable e is dependent on the unknown system parameters $a_i, i = 0, \cdots, n-1$ .It follows that the following equality is valid

$$\alpha_{ab}(t) y_M(t) = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i)}(t,\tau) y_M(\tau) d\tau + e_1(t) \tag{4.45}$$

$$\int_a^b \alpha_{ab}(\tau) y_M(\tau) d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K^1_{DS(i)}(t,\tau) y_M(\tau) d\tau + e_2(t) \tag{4.46}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau) y_M(\tau) d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K^2_{DS(i)}(t,\tau) y_M(\tau) d\tau + e_3(t) \tag{4.47}$$

$$\int_a^b \alpha_{ab}(\tau)(t-\tau)^2 y_M(\tau) d\tau = \sum_{i=0}^{n} \beta_i \int_a^b K^3_{DS(i)}(t,\tau) y_M(\tau) d\tau + e_4(t) \tag{4.48}$$

which can be further written in matrix form as

$$\begin{bmatrix} \alpha_{ab}(t)y_M(t) - \beta_n \int_a^b K_{DS(n)}(t,\tau)y_M(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)y_M(\tau)d\tau - \beta_n \int_a^b K^1_{DS(n)}(t,\tau)y_M(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t-\tau)y_M(\tau)d\tau - \beta_n \int_a^b K^2_{DS(n)}(t,\tau)y_M(\tau)d\tau \\ \int_a^b \alpha_{ab}(\tau)(t-\tau)^2 y_M(\tau)d\tau - \beta_n \int_a^b K^3_{DS(n)}(t,\tau)y_M(\tau)d\tau \end{bmatrix} = \begin{bmatrix} \int_a^b K_{DS(k)}(t,\tau)y_M(\tau)d\tau \\ \int_a^b K^1_{DS(k)}(t,\tau)y_M(\tau)d\tau \\ \int_a^b K^2_{DS(k)}(t,\tau)y_M(\tau)d\tau \\ \int_a^b K^3_{DS(k)}(t,\tau)y_M(\tau)d\tau \end{bmatrix} \overline{a} + \begin{bmatrix} e1 \\ e2 \\ e3 \\ e4 \end{bmatrix} \tag{4.49}$$

and has the random regressor matrix

$$\begin{bmatrix} \int_a^b K_{DS(0),y}(t,\tau)y_M(\tau)d\tau, \cdots, \int_a^b K_{DS(n-1),y}(t,\tau)y_M(\tau)d\tau \\ \int_a^b K^1_{DS(0),y}(t,\tau)y_M(\tau)d\tau, \cdots, \int_a^b K^1_{DS(n-1),y}(t,\tau)y_M(\tau)d\tau \\ \int_a^b K^2_{DS(0),y}(t,\tau)y_M(\tau)d\tau, \cdots, \int_a^b K^2_{DS(n-1),y}(t,\tau)y_M(\tau)d\tau \\ \int_a^b K^3_{DS(0),y}(t,\tau)y_M(\tau)d\tau, \cdots, \int_a^b K^3_{DS(n-1),y}(t,\tau)y_M(\tau)d\tau \end{bmatrix}^T \tag{4.50}$$

with corresponding error terms

$$e1 = \alpha_{ab}(t)y_M(t) - \int_a^b K_{DS}(t,\tau)y_M(\tau)d\tau \tag{4.51}$$

$$e2 = \int_a^b \alpha_{ab}(\tau)y_M(\tau)d\tau - \beta \int_a^b K^1_{DS}(t,\tau)y_M(\tau)d\tau \tag{4.52}$$

$$e3 = \int_a^b \alpha_{ab}(\tau)(t-\tau)y_M(\tau)d\tau - \int_a^b K^2_{DS}(t,\tau)y_M(\tau)d\tau \tag{4.53}$$

$$e4 = \int_a^b \alpha_{ab}(\tau)(t-\tau)^2 y_M(\tau)d\tau - \int_a^b K^3_{DS}(t,\tau)y_M(\tau)d\tau \tag{4.54}$$

On observing the above error terms it can be seen that assumptions of Gauss-Markov Theorem are violated and additionally the error fails to be homoskedastic.

4.2 Heteroskedasticity [12]

The existence of heteroskedasticity is a major concern in regression analysis, Heteroskedasticity has serious consequences for the Ordinary Least square estimator. Despite that the fact that the OLS estimator remains unbiased, the estimated regression error is wrong while confidence intervals cannot be relied on.In the presence of heteroskedastic errors, regression using Feasible Generalized Least Squares (FGLS) offers potential efficiency gains over Ordinary Least Squares (OLS). Feasible generalized least squares (FGLS) employs inverse covariance weighting in the regression error minimization problem associated with equations (4.45),(4.46),(4.47) and (4.48). Let $Q(y_M)$ and $P(y_M)$ as defined in the (4.36) be the matrices corresponding to N samples of the measurement process realization y_M at a batch of knots $t_1, \cdots, t_N$. Then the stochastic regression error vector is given by

$$e_n := [e_n(t_i), \cdots, e_n(t_N)]^T = Q(y_M) - P(y_M)\bar{a} \tag{4.55}$$

where $e_n(t_i)$ are as in (4.51), (4.52),(4.53) and (4.54). The standard GLS regression error minimization for estimation of the parameter vector $\bar{a}$ is

$$\min_{\bar{a}} \left([Q(y_M) - P(y_M)\bar{a}]^T S [Q(y_M) - P(y_M)\bar{a})] \right) \tag{4.56}$$

$$\text{with} \quad S := [\mathrm{Cov}(e)]^{-1} \tag{4.57}$$

Applying the expectation operator to equations (4.51),(4.52),(4.53) and (4.54) and using the properties of the Gaussian noise yields

$$\mathrm{E}[e_1(t)] = \mathrm{E}[\alpha_{ab}(t)(\sigma\dot{W}(t))] - \mathrm{E}[\int_a^b K_{DS,y}(t,\tau)\sigma\dot{W}(\tau)] = 0 \tag{4.58}$$

$$\mathrm{E}[e_2(t)] = \mathrm{E}[\int_a^b \alpha_{ab}(\tau)(\sigma\dot{W}(\tau))] - \mathrm{E}[\int_a^b K^1_{DS,y}(t,\tau)\sigma\dot{W}(\tau)d\tau] = 0 \tag{4.59}$$

$$\mathrm{E}[e_3(t)] = \mathrm{E}[\int_a^b \alpha_{ab}(\tau)(t-\tau)(\sigma\dot{W}(\tau))] - \mathrm{E}[\int_a^b K^2_{DS,y}(t,\tau)\sigma\dot{W}(\tau)d\tau] = 0 \tag{4.60}$$

$$\mathrm{E}[e_4(t)] = \mathrm{E}[\int_a^b \alpha_{ab}(\tau)(t-\tau)(\sigma\dot{W}(\tau))] - \mathrm{E}[\int_a^b K^3_{DS,y}(t,\tau)\sigma\dot{W}(\tau)d\tau] = 0 \tag{4.61}$$

4.3 Error Covariance Matrix

The following facts and properties of Gaussian White noise are recalled for the calculation of Covariance.The generalized derivative $\dot{W}$ of Wiener process W is defined by the following equality that needs to hold for all smooth functions g with compact support,

$$g(t)W(t) = \int_0^t g(s)\dot{W}(s)ds + \int_0^t \dot{g}(s)W(s)ds; \quad t \in [a, b] \tag{4.62}$$

The integral $\int_0^t \dot{g}(s)W(s)ds$ is well defined for any square integrable function g; i.e. g does not have to be smooth, but

$$\int_0^t g(s)^2 ds < \infty \tag{4.63}$$

The generalized expectation and covariance functions of white noise are given by:

$$E[\dot{W}_t] \equiv 0 \tag{4.64}$$

$$Cov[\dot{W}_t, \dot{W}_s] = E[\dot{W}_t\dot{W}_s] = \delta(t - s) \tag{4.65}$$

$$Var[\dot{W}_t] = E[\dot{W}_t^2] = 1 \tag{4.66}$$

where is the delta Dirac distribution (strictly: a linear functional defined on the space of tempered distributions) but acting on square integrable functions as an evaluation functional:

$$\int_a^b g(s)\delta(t - s)ds = g(t), \quad t \in [a, b] \tag{4.67}$$

Using the properties (4.68) - (4.73) of white noise and recalling that the kernel functions are Hilbert-Schmidt, hence are square integrable, the covariance calculation is then carried out as follows

$$
\begin{aligned}
&Cov[e_1(t_i), e_1(t_j)] = E[e_1(t_i)e_1(t_j)] \\
&= E\Bigg[\Big[\alpha_{ab}(t_i)\sigma\dot{W}(t_i) - \int_a^b K_{DS}(t_i,s)\sigma\dot{W}(s)ds\Big] \\
&\qquad\qquad\qquad \Big[\alpha_{ab}(t_j)\sigma\dot{W}(t_j) - \int_a^b K_{DS}(t_j,\tau)\sigma\dot{W}(\tau)d\tau\Big]\Bigg] \\
&= \Bigg[E\Big[\sigma^2\alpha_{ab}(t_i)\alpha_{ab}(t_j)\dot{W}(t_i)\dot{W}(t_j)\Big] \\
&\qquad - E\Big[\sigma^2\alpha_{ab}(t_i)\dot{W}(t_i)\int_a^b K_{DS}(t_j,\tau)\dot{W}(\tau)d\tau\Big] \\
&\qquad\quad - E\Big[\sigma^2\alpha_{ab}(t_j)\dot{W}(t_j)\int_a^b K_{DS}(t_i,s)\dot{W}(s)ds\Big] \\
&\qquad\qquad + E\Big[\sigma^2\int_a^b\int_a^b K_{DS}(t_i,s)K_{DS}(t_j,\tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big]\Bigg] \\
&= E\Big[\sigma^2\alpha_{ab}(t_i)\alpha_{ab}(t_j)\dot{W}(t_i)\dot{W}(t_j)\Big] \\
&\qquad - E\Big[\sigma^2\alpha_{ab}(t_i)\int_a^b K_{DS}(t_j,\tau)\dot{W}(t_i)\dot{W}(\tau)d\tau\Big] \\
&\qquad\quad - E\Big[\sigma^2\alpha_{ab}(t_j)\int_a^b K_{DS}(t_i,s)\dot{W}(t_j)\dot{W}(s)ds\Big] \\
&\qquad\qquad + E\Big[\sigma^2\int_a^b\int_a^b K_{DS}(t_i,s)K_{DS}(t_j,\tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big] \\
&= \sigma^2\alpha_{ab}(t_i)\alpha_{ab}(t_j)E\Big[\dot{W}(t_i)\dot{W}(t_j)\Big] \\
&\qquad - \sigma^2\alpha_{ab}(t_i)\int_a^b K_{DS}(t_j,\tau)E\Big[\dot{W}(t_i)\dot{W}(\tau)\Big]d\tau \\
&\qquad\quad - \sigma^2\alpha_{ab}(t_j)\int_a^b K_{DS}(t_i,s)E\Big[\dot{W}(t_j)\dot{W}(s)\Big]ds \\
&\qquad\qquad + \sigma^2\int_a^b\int_a^b K_{DS}(t_i,s)K_{DS}(t_j,\tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau \\
&Cov[e_1(t_i), e_1(t_j)] = \sigma^2\alpha_{ab}(t_i)\alpha_{ab}(t_j)\delta(t_i-t_j) - \sigma^2\alpha_{ab}(t_i)K_{DS}(t_i,t_j) - \sigma^2\alpha_{ab}(t_j)K_{DS}(t_j,t_i) \\
&\qquad\quad + \sigma^2\int_a^b K_{DS}(t_i,s)K_{DS}(t_j,s)ds
\end{aligned}
\tag{4.68}
$$

$$
\begin{aligned}
Cov[e_2(t_i), e_2(t_j)] &= E[e_2(t_i)e_2(t_j)] \\
&= E\Bigg[\Big[\int_a^b \alpha_{ab}(s)\sigma\dot{W}(s)ds - \int_a^b K_{DS}^1(t_i,s)\sigma\dot{W}(s)ds\Big] \\
&\qquad \Big[\int_a^b \alpha_{ab}(\tau)\sigma\dot{W}(\tau)d\tau - \int_a^b K_{DS}^1(t_j,\tau)\sigma\dot{W}(\tau)d\tau\Big]\Bigg] \\
&= \Bigg[E\Big[\sigma^2\int_a^b\int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big] \\
&\qquad - E\Big[\sigma^2\int_a^b\int_a^b \alpha_{ab}(s)\dot{W}(s)K_{DS}^1(t_j,\tau)\dot{W}(\tau)dsd\tau\Big] \\
&\qquad - E\Big[\sigma^2\int_a^b\int_a^b \alpha_{ab}(\tau)\dot{W}(\tau)K_{DS}^1(t_i,s)\dot{W}(s)dsd\tau\Big] \\
&\qquad + E\Big[\sigma^2\int_a^b\int_a^b K_{DS}^1(t_i,s)K_{DS}^1(t_j,\tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big]\Bigg] \\
&= \Bigg[\sigma^2\int_a^b\int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau - \\
&\qquad \sigma^2\int_a^b\int_a^b \alpha_{ab}(s)K_{DS}^1(t_j,\tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau \\
&\qquad - \sigma^2\int_a^b\int_a^b \alpha_{ab}(\tau)K_{DS}^1(t_i,s)E\Big[\dot{W}(\tau)\dot{W}(s)\Big]dsd\tau \\
&\qquad + \sigma^2\int_a^b\int_a^b K_{DS}^1(t_i,s)K_{DS}^1(t_j,\tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau\Bigg] \\
&= \Bigg[\sigma^2\int_a^b\int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)\delta(s-\tau)dsd\tau \\
&\qquad - \sigma^2\int_a^b\int_a^b \alpha_{ab}(s)K_{DS}^1(t_j,\tau)\delta(s-\tau)dsd\tau \\
&\qquad - \sigma^2\int_a^b\int_a^b \alpha_{ab}(\tau)K_{DS}^1(t_i,s)\delta(s-\tau)dsd\tau \\
&\qquad + \sigma^2\int_a^b\int_a^b K_{DS}^1(t_i,s)K_{DS}^1(t_j,\tau)\delta(s-\tau)dsd\tau\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
Cov[e_2(t_i), e_2(t_j)] = \sigma^2\Bigg[&\int_a^b \alpha_{ab}(s)\alpha_{ab}(s)ds - \int_a^b \alpha_{ab}(s)K_{DS}^1(t_j,s)ds - \int_a^b \alpha_{ab}(s)K_{DS}^1(t_i,s)ds \\
&+ \int_a^b K_{DS}^1(t_i,s)K_{DS}^1(t_j,s)ds\Bigg] \qquad (4.69)
\end{aligned}
$$

$$
\begin{aligned}
&Cov[e_3(t_i), e_3(t_j)] = E[e_3(t_i)e_3(t_j)] \\
&= E\Big[\Big[\int_a^b \alpha_{ab}(s)(t_i - s)\sigma\dot{W}(s)ds - \int_a^b K_{DS}^2(t_i, s)\sigma\dot{W}(s)ds\Big] \\
&\qquad\qquad \Big[\int_a^b \alpha_{ab}(\tau)(t_j - \tau)\sigma\dot{W}(\tau)d\tau - \int_a^b K_{DS}^2(t_j, \tau)\sigma\dot{W}(\tau)d\tau\Big]\Big] \\
&= \Big[E\Big[\sigma^2 \int_a^b \int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)(t_i - s)(t_j - \tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big] \\
&\qquad - E\Big[\sigma^2 \int_a^b \int_a^b \alpha_{ab}(s)(t_i - s)\dot{W}(s)K_{DS}^2(t_j, \tau)\dot{W}(\tau)dsd\tau\Big] \\
&\qquad\qquad - E\Big[\sigma^2 \int_a^b \int_a^b \alpha_{ab}(\tau)(t_j - \tau)\dot{W}(\tau)K_{DS}^2(t_i, s)\dot{W}(s)dsd\tau\Big] \\
&\qquad\qquad\qquad + E\Big[\sigma^2 \int_a^b \int_a^b K_{DS}^2(t_i, s)K_{DS}^2(t_j, \tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big]\Big] \\
&= \Big[\sigma^2 \int_a^b \int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)(t_i - s)(t_j - \tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau \\
&\qquad - \sigma^2 \int_a^b \int_a^b \alpha_{ab}(s)(t_i - s)K_{DS}^2(t_j, \tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau \\
&\qquad\qquad - \sigma^2 \int_a^b \int_a^b \alpha_{ab}(\tau)(t_j - \tau)K_{DS}^2(t_i, s)E\Big[\dot{W}(\tau)\dot{W}(s)\Big]dsd\tau \\
&\qquad\qquad\qquad + \sigma^2 \int_a^b \int_a^b K_{DS}^2(t_i, s)K_{DS}^2(t_j, \tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau\Big] \\
&= \Big[\sigma^2 \int_a^b \int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)(t_i - s)(t_j - \tau)\delta(s - \tau)dsd\tau \\
&\qquad - \sigma^2 \int_a^b \int_a^b \alpha_{ab}(s)(t_i - s)K_{DS}^2(t_j, \tau)\delta(s - \tau)dsd\tau \\
&\qquad\qquad - \sigma^2 \int_a^b \int_a^b \alpha_{ab}(\tau)(t_j - \tau)K_{DS}^2(t_i, s)\delta(s - \tau)dsd\tau \\
&\qquad\qquad\qquad + \sigma^2 \int_a^b \int_a^b K_{DS}^2(t_i, s)K_{DS}^2(t_j, \tau)\delta(s - \tau)dsd\tau\Big] \\
&Cov[e_3(t_i), e_3(t_j)] = \sigma^2\Big[\int_a^b \alpha_{ab}(s)\alpha_{ab}(s)(t_i - s)(t_j - s)ds - \int_a^b \alpha_{ab}(s)(t_i - s)K_{DS}^2(t_j, s)ds \\
&\qquad\qquad - \int_a^b \alpha_{ab}(s)(t_j - s)K_{DS}^2(t_i, s)ds + \int_a^b K_{DS}^2(t_i, s)K_{DS}^2(t_j, s)ds\Big]
\end{aligned}
\tag{4.70}
$$

$$
\begin{aligned}
&Cov[e_4(t_i), e_4(t_j)] = E[e_4(t_i)e_4(t_j)] \\
&= E\Big[\Big[\int_a^b \frac{1}{2}\alpha_{ab}(s)(t_i-s)^2\sigma\dot{W}(s)ds - \int_a^b K_{DS}^3(t_i,s)\sigma\dot{W}(s)ds\Big] \\
&\qquad \Big[\int_a^b \frac{1}{2}\alpha_{ab}(\tau)(t_j-\tau)^2\sigma\dot{W}(\tau)d\tau - \int_a^b K_{DS}^3(t_j,\tau)\sigma\dot{W}(\tau)d\tau\Big]\Big] \\
&= E\Big[\sigma^2\frac{1}{4}\int_a^b\int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)(t_i-s)^2(t_j-\tau)^2\dot{W}(s)\dot{W}(\tau)dsd\tau\Big] \\
&\quad - E\Big[\sigma^2\frac{1}{2}\int_a^b\int_a^b \alpha_{ab}(s)(t_i-s)^2\dot{W}(s)K_{DS}^3(t_j,\tau)\dot{W}(\tau)dsd\tau\Big] \\
&\quad - E\Big[\sigma^2\frac{1}{2}\int_a^b\int_a^b \alpha_{ab}(\tau)(t_j-\tau)^2\dot{W}(\tau)K_{DS}^3(t_i,s)\dot{W}(s)dsd\tau\Big] \\
&\quad + E\Big[\sigma^2\int_a^b\int_a^b K_{DS}^3(t_i,s)K_{DS}^3(t_j,\tau)\dot{W}(s)\dot{W}(\tau)dsd\tau\Big] \\
&= \Bigg[\sigma^2\frac{1}{4}\int_a^b\int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)(t_i-s)^2(t_j-\tau)^2E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau \\
&\quad - \sigma^2\frac{1}{2}\int_a^b\int_a^b \alpha_{ab}(s)(t_i-s)^2K_{DS}^3(t_j,\tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau \\
&\quad - \sigma^2\frac{1}{2}\int_a^b\int_a^b \alpha_{ab}(\tau)(t_j-\tau)^2K_{DS}^3(t_i,s)E\Big[\dot{W}(\tau)\dot{W}(s)\Big]dsd\tau \\
&\quad + \sigma^2\int_a^b\int_a^b K_{DS}^3(t_i,s)K_{DS}^3(t_j,\tau)E\Big[\dot{W}(s)\dot{W}(\tau)\Big]dsd\tau\Bigg] \\
&= \Bigg[\sigma^2\frac{1}{4}\int_a^b\int_a^b \alpha_{ab}(s)\alpha_{ab}(\tau)(t_i-s)^2(t_j-\tau)^2\delta(s-\tau)dsd\tau \\
&\quad - \sigma^2\frac{1}{2}\int_a^b\int_a^b \alpha_{ab}(s)(t_i-s)^2K_{DS}^3(t_j,\tau)\delta(s-\tau)dsd\tau \\
&\quad - \sigma^2\frac{1}{2}\int_a^b\int_a^b \alpha_{ab}(\tau)(t_j-\tau)^2K_{DS}^3(t_i,s)\delta(s-\tau)dsd\tau \\
&\quad + \sigma^2\int_a^b\int_a^b K_{DS}^3(t_i,s)K_{DS}^3(t_j,\tau)\delta(s-\tau)dsd\tau\Bigg] \\
&Cov[e_4(t_i), e_4(t_j)] = \sigma^2\Bigg[\frac{1}{4}\int_a^b \alpha_{ab}(s)\alpha_{ab}(s)(t_i-s)^2(t_j-s)^2ds - \frac{1}{2}\int_a^b \alpha_{ab}(s)(t_i-s)^2K_{DS}^3(t_j,s)ds \\
&\qquad - \frac{1}{2}\int_a^b \alpha_{ab}(s)(t_j-s)^2K_{DS}^3(t_i,s)ds + \int_a^b K_{DS}^3(t_i,s)K_{DS}^3(t_j,s)ds\Bigg]
\end{aligned}
\tag{4.71}
$$

Covariance Matrix(S^{-1}) =

$$\begin{bmatrix} Cov[e_1(t_i), e_1(t_j)] & \dots & \dots & 0 \\ \vdots & Cov[e_2(t_i), e_2(t_j)] & \dots & \vdots \\ \vdots & \dots & Cov[e_3(t_i), e_3(t_j)] & \vdots \\ 0 & \dots & \dots & Cov[e_4(t_i), e_4(t_j)] \end{bmatrix} \tag{4.72}$$

4.4 FGLS algorithm

"The covariance matrix depends on the unknown variance σ^2 and also on the unknown parameter vector $\bar{a}$ in the K_{DS} kernels.Therefore the Standard OLS estimator cannot be applied directly.Hence we use a Recursive Feasible GLS in which the covariance matrix is estimated progressively. Letting $Q_i - P_i\bar{a}$ denote the regression error e_i in batch i, the FGLS algorithm computes

$$\hat{a}_k = \arg\min_{\bar{a}} \left(\sum_{i=1}^{k} (Q_i - P_i\bar{a})^T S_i (Q_i - P_i\bar{a}) \right) \tag{4.73}$$

where $\hat{a}_k$ is the parameter estimate update at iteration k of the algorithm. Each weighting matrix S_{k+1}, is calculated as the inverse of the covariance matrix based on the parameter estimate $\hat{a}_k$ and an estimate of the variance σ^2 obtained from the residual trajectory $y_M(t) - y_E(t)$ in previous iteration k, where y_E signifies the estimated/reconstructed output" [12]. The FGLS algorithm is generally a modified version of standard recursive least squares algorithm [6].The FGLS employs an Inverse covarinace weighting with a recursive approach is presented below.

At iteration $k + 1$, the algorithm strives to minimize

$$\begin{aligned} &\min(\bar{e}_{k+1}^T \bar{S}_{k+1} \bar{e}_{k+1}) \\ \text{subject to:} \quad &\bar{Q}_{k+1} = \bar{P}_{k+1}\bar{a}_{k+1} + \bar{e}_{k+1} \end{aligned} \tag{4.74}$$

where

$$\bar{Q}_{k+1} = \begin{bmatrix} Q_0 \\ Q_1 \\ \vdots \\ Q_{k+1} \end{bmatrix}; \quad \bar{P}_{k+1} = \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_{k+1} \end{bmatrix}; \quad \bar{e} = \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_{k+1} \end{bmatrix}$$

and

$$\bar{S}_{k+1} = \text{diag}(S_0, S_1, \ldots, S_{k+1}) \tag{4.75}$$

The solution of the above is written as

$$(\bar{P}^T_{k+1}\bar{S}_{k+1}\bar{P}_{k+1})\hat{a}_{k+1} = \bar{P}^T_{k+1}\bar{S}_{k+1}\bar{Q}_{k+1} \tag{4.76}$$

or in summation form as

$$\left(\sum_{i=0}^{k+1} P_i^T S_i P_i\right)\hat{a}_{k+1} = \sum_{i=0}^{k+1} P_i^T S_i Q_i \tag{4.77}$$

Defining

$$M_{k+1} = \sum_{i=0}^{k+1} P_i^T S_i P_i \tag{4.78}$$

the recursion for M_{k+1} is:

$$M_{k+1} = M_k + P^T_{k+1} S_{k+1} P_{k+1} \tag{4.79}$$

Rearranging (4.77) gives

$$\begin{aligned} \hat{a}_{k+1} &= M^{-1}_{k+1}\left[\left(\sum_{i=0}^{k} P_i^T S_i P_i\right)\hat{a}_k + P^T_{k+1} S_{k+1} Q_{k+1}\right] \\ &= M^{-1}_{k+1}\left[M_k \hat{a}_k + P^T_{k+1} S_{k+1} Q_{k+1}\right] \end{aligned} \tag{4.80}$$

Another form of 4.80 is delivered by the recursion 4.79 and reads

$$\begin{aligned} \hat{a}_{k+1} &= \hat{a}_k - M^{-1}_{k+1}(P^T_{k+1} S_{k+1} P_{k+1}\hat{a}_k - P^T_{k+1} S_{k+1} Q_{k+1}) \\ &= \hat{a}_k + M^{-1}_{k+1} P^T_{k+1} S_{k+1}(Q_{k+1} - P_{k+1}\hat{a}_k) \end{aligned} \tag{4.81}$$

A recursion for M_{k+1}^{-1} is obtained by applying the following identity to the recursion in 4.79

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1} \tag{4.82}$$

which yields

$$M_{k+1}^{-1} = M_k^{-1} - M_k^{-1}P_{k+1}^T(P_{k+1}M_k^{-1}P_{k+1}^T + S_{k+1}^{-1})^{-1}P_{k+1}M_k^{-1}$$

Defining $R_{k+1} = M_{k+1}^{-1}$ the latter becomes

$$R_{k+1} = R_k - R_kP_{k+1}^T(S_{k+1}^{-1} + P_{k+1}R_kP_{k+1}^T)^{-1}P_{k+1}R_k \tag{4.83}$$

Equations 4.82 and 4.83 with inverse covariance weighting constitute the Feasible Generalised least squares algorithm.

The initial estimate of σ^2 is calculated as the *empirical* variance of $(y_M - y_E)$ where y_E is the estimated output corresponding to the parameter values obtained in iteration $k = 0$. This is updated at each consecutive iteration by the same empirical method.
To obtain an initial estimate for the weighting matrix, S_0, the reproducing property of Theorem 1 is first differentiated three times to deliver three additional linearly independent regression equations, this to match the number of the parameters in $\bar{a}$ with the number of regression equations. An ordinary OLS is next used to deliver crude estimates of $\hat{a}_0$ that can serve the evaluation of the covariance matrix in S_0.

4.5 Reconstruction of the output derivatives

Given a measurement process realization $\overline{y_M}$ on $[a, b]$, the derivatives can be reconstructed using,

$$y^{(i)}(t) = \int_a^b K_{DS}^i(t, \tau)\hat{y}(\tau)\mathrm{d}\tau \tag{4.84}$$

where, K_{DS}^i are the kernel representation for output derivatives. In this thesis we consider $i = 1$,2 and 3 . The formulae for kernel representation of output derivatives are developed in Chapter 2.

Chapter 5

Results and Discussion

We employed the Feasible Generalised Least Squares algorithm for parameter and state estimation in Chapter 4. Below we provide examples of fourth order systems and consider different noise perturbations in the measured output.We have used this specific example so as to compare the results with [12]

Example1: [12]

Consider a fourth order system as described below:

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -5 & -5 & 0 \end{bmatrix} x \; ; y = x_1 \; ; \; x(0) = [0, 0, 0, 1] \tag{5.1}$$

with its corresponding characteristic equation

$$y^{(4)}(t) + a_3 y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{5.2}$$

The parameters a_0, a_1, a_2,a_3 are assumed to be unknown.

The measured realization of the output y_M is obtained as $y_M = y_T + \eta^{\sigma}$ with variance σ and sampled as needed.As the quality of the GLS algorithm in [12] is known to rely on large sample theory, the FGLS with the multiple regression as explained in chapter4 does not rely on large sample theory when compared to [12] .when the system is subjected to low noise and moderate noise levels batches of $N = 500$, knots are sampled from a uniform distribution over $[a, b] = [0, 6]$.However, when the system is subject to high noise levels the sample size in increased to achieve closer estimates.The threshold value for stopping the algorithm is $\epsilon = 0.01$.The estimated system parameters with high noise level were found as they are presented in Table (5.1) with the number of sample points $N = 2000$ which were selected to be equidistant in $[a, b]$.

	a_0	a_1	a_2	a_3
True values	1	5	5	0
Estimated Values	1.71	3.6848	5.904,	-0.2215

Table 5.1 Table showing true and estimated parameter values from a true output with AWGN $\mu = 0$ and $\sigma = 1.75$ and N=2000

Once the parameter estimates are obtained, we reconstruct the output and its derivatives as mentioned in section 4.5 chapter 4.

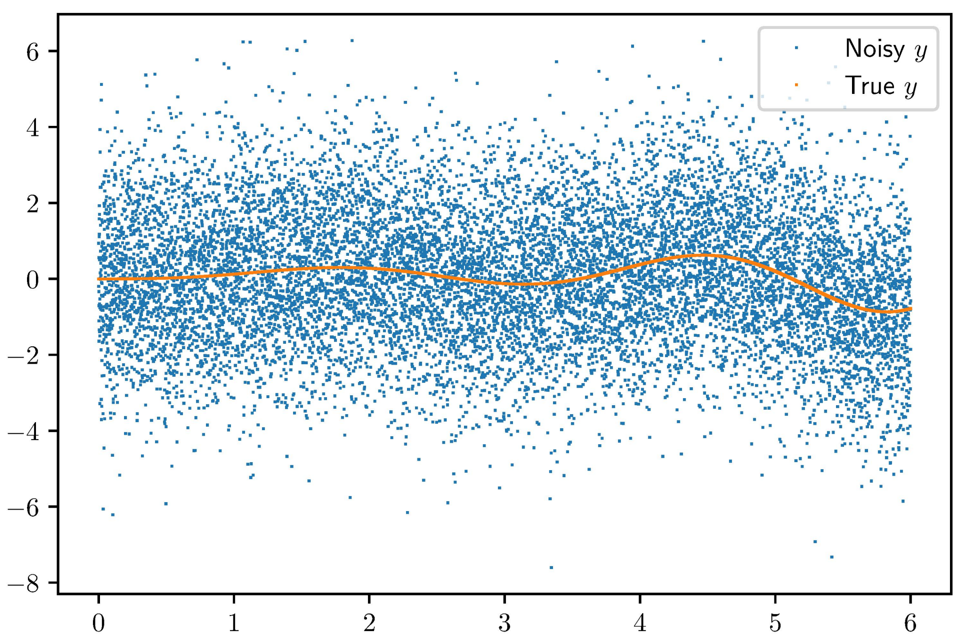

Figure 5.1 True and noisy system outputs with AWGN of $\mu = 0$ and $\sigma =$ 1.75 and N=2000

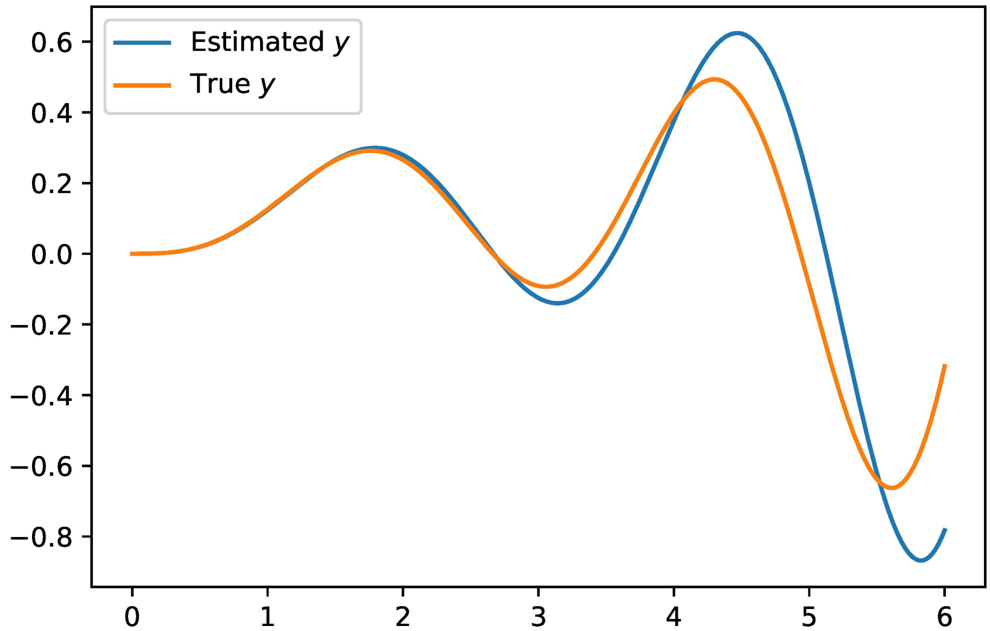

Figure 5.2 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 1.75$

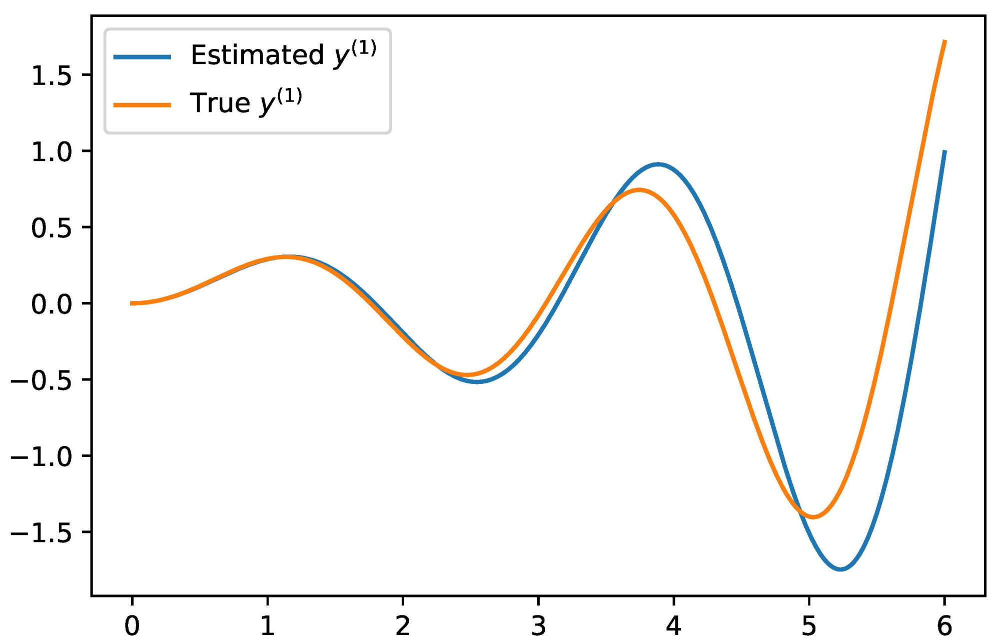

Figure 5.3 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.75$

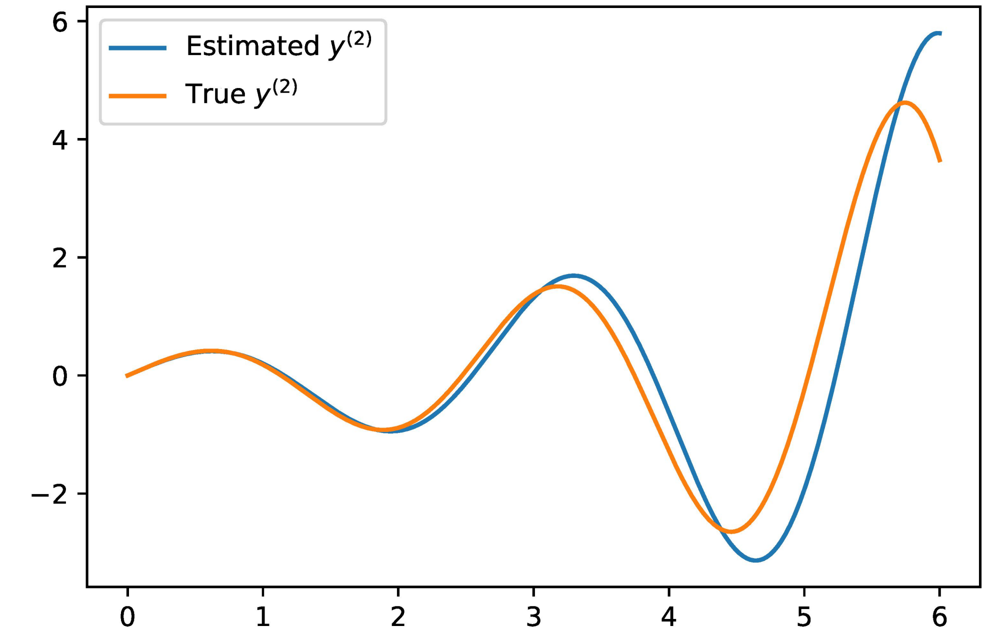

Figure 5.4 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.75$

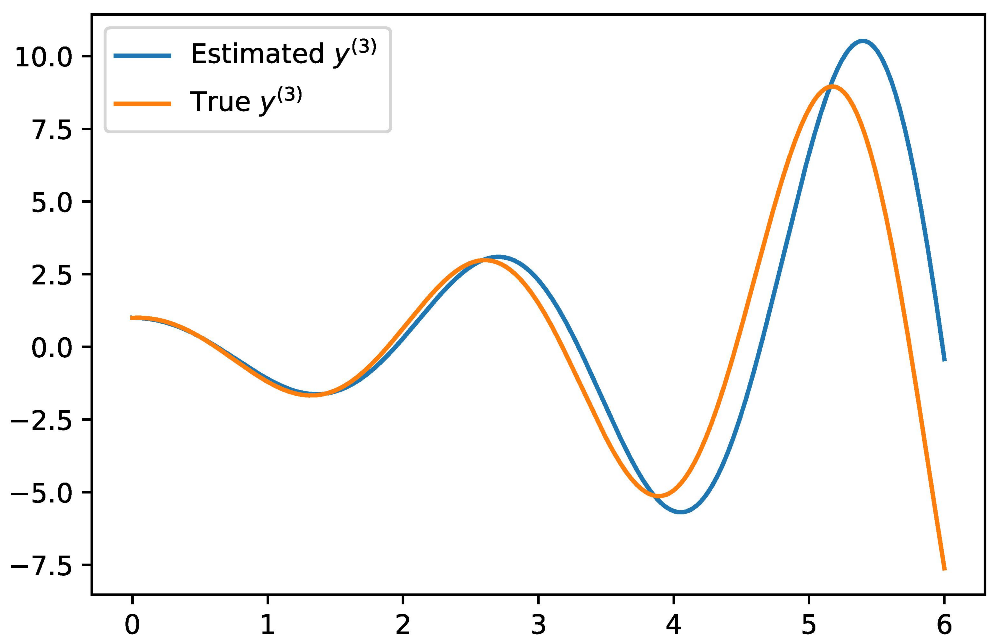

Figure 5.5 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.75$

5.1 Study of the influence of measurement noise levels and number of sample points on estimation

Consider again the system in (5.6), on the same time interval $[0, 6]$. Three different noise levels are employed to emulate noise-perturbed output measurement in (5.6). Simultaneously, the number of sample points (N) is varied to study its influence on the accuracy of estimation. Additionally, to better understand how the number of sample points (N) affects the accuracy of estimation we compute the root mean square deviation, RMSD, as:

$$RMSD = \sqrt{\sum_{i=1}^{N} \frac{1}{n} [y_T(t_i) - \hat{y}(t_i)]^2} \tag{5.3}$$

where $y_T(t_i)$ and $\hat{y}(t_i)$ are: the true system output and the estimated output at time instant t_i, respectively.

Table (5.2.2) shows how the parameter estimation accuracy varies with increasing noise levels and number of sample points N.

Variance	Samples(N)	$a0$	$a1$	$a2$	$a3$
0.25	500	1.0130	5.0264	5.0100	0.0080
0.5	500	0.9849	4.978,	5.0810	0.0090
0.75	500	1.0080	4.9212	4.9980	-0.0660
1.00	500	0.9800	4.9600	4.9100	-0.0900
1.25	500	1.210	5.1091	4.8944	-0.1248
	2000	1.0516	4.9830	4.8964	-0.1226
1.5	500	1.3289	4.9148	5.1204	0.1315
	2000	1.1289	4.9762	5.3291	-0.03
1.75	2000	1.71	3.6848	5.904	-0.2215

Table 5.2 Estimates of parameter values and $RMSD$ for various noise levels and sample size N

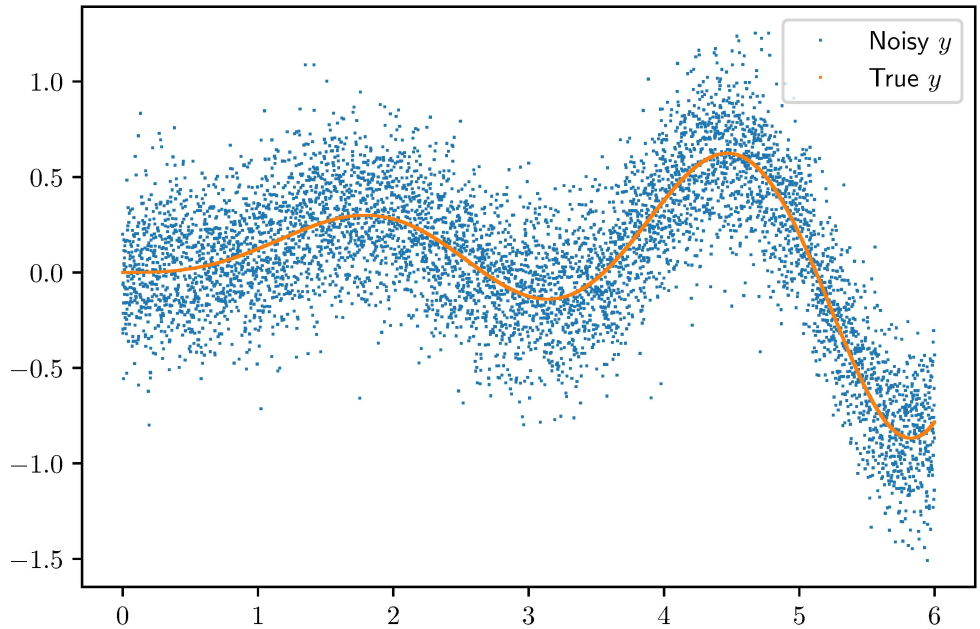

Figure 5.6 True and noisy system outputs with AWGN of $\mu = 0$ and $\sigma = 0.25$ and N=500

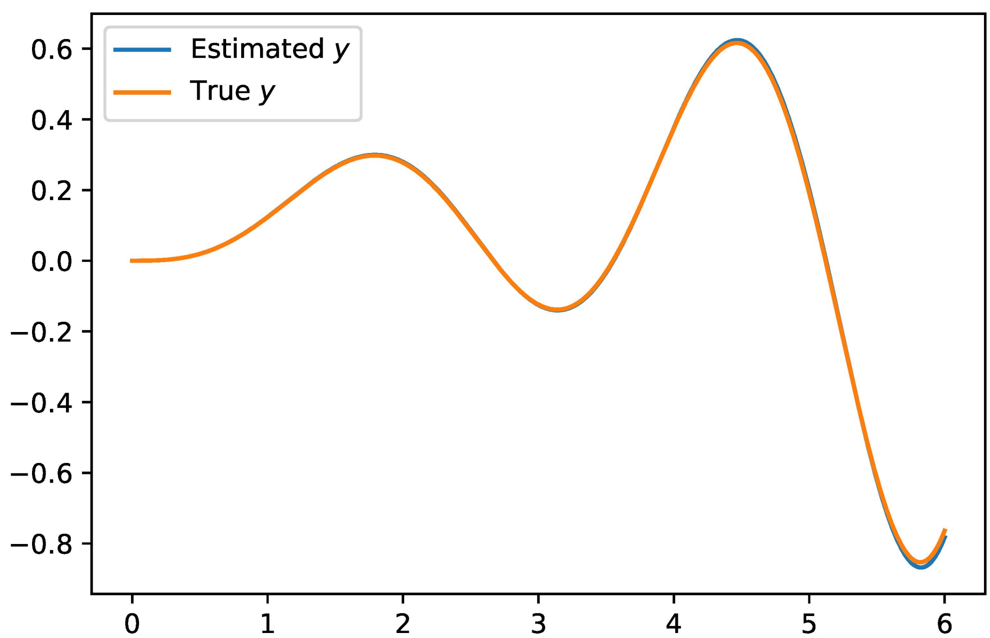

Figure 5.7 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 0.25$

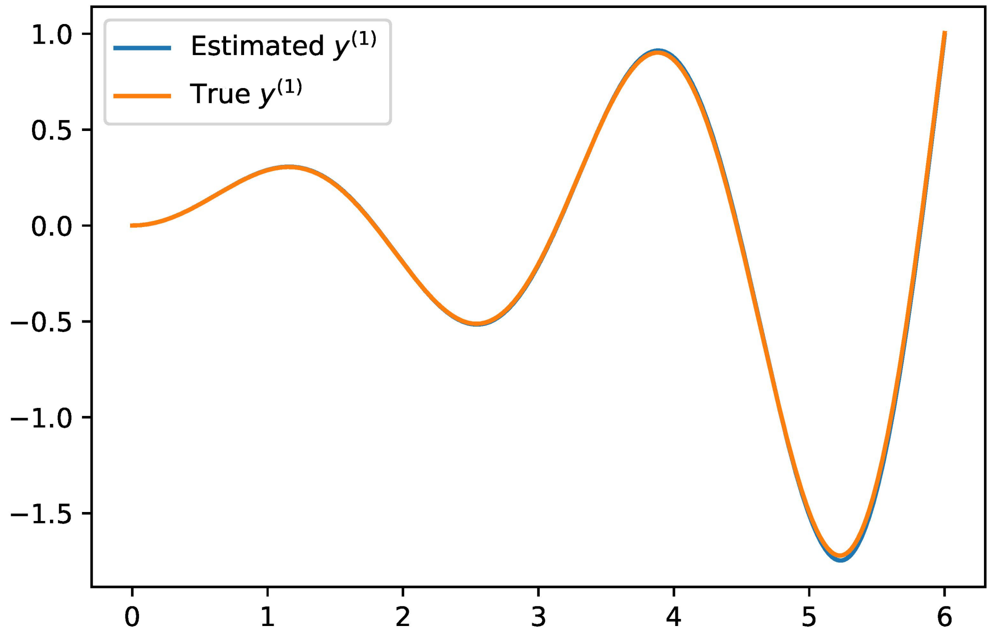

Figure 5.8 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.25$

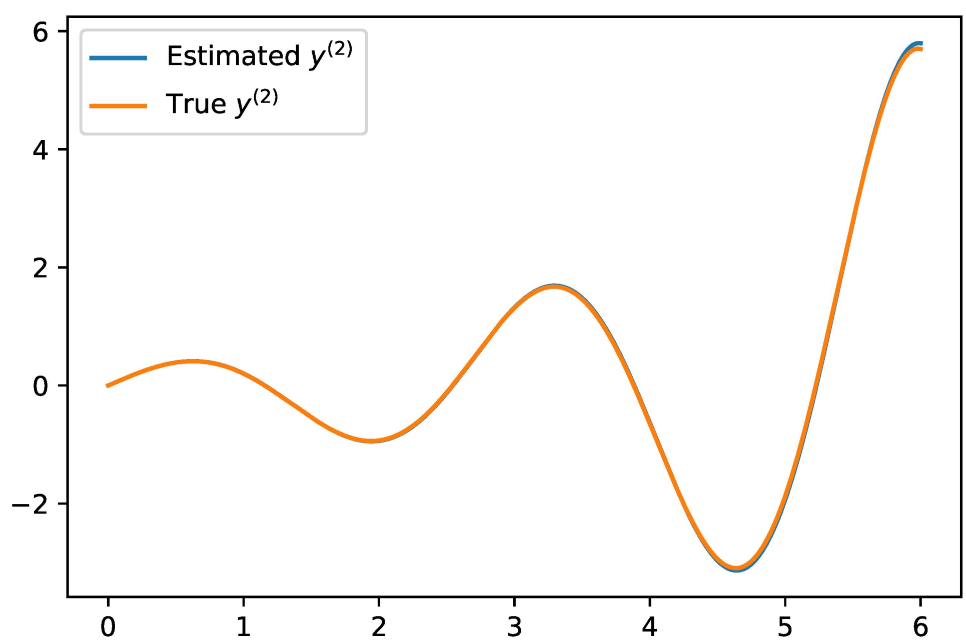

Figure 5.9 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.25$

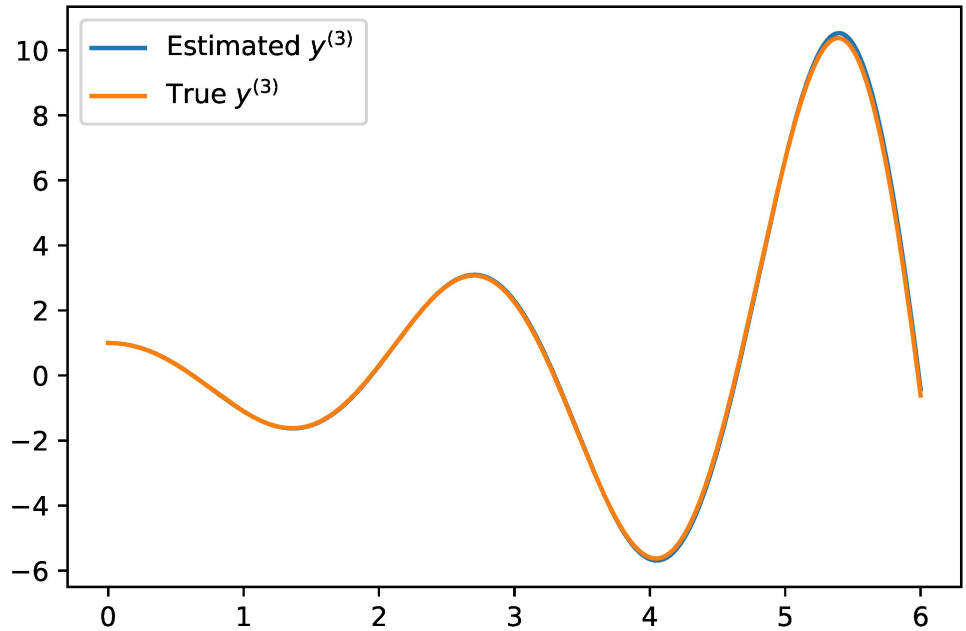

Figure 5.10 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.25$

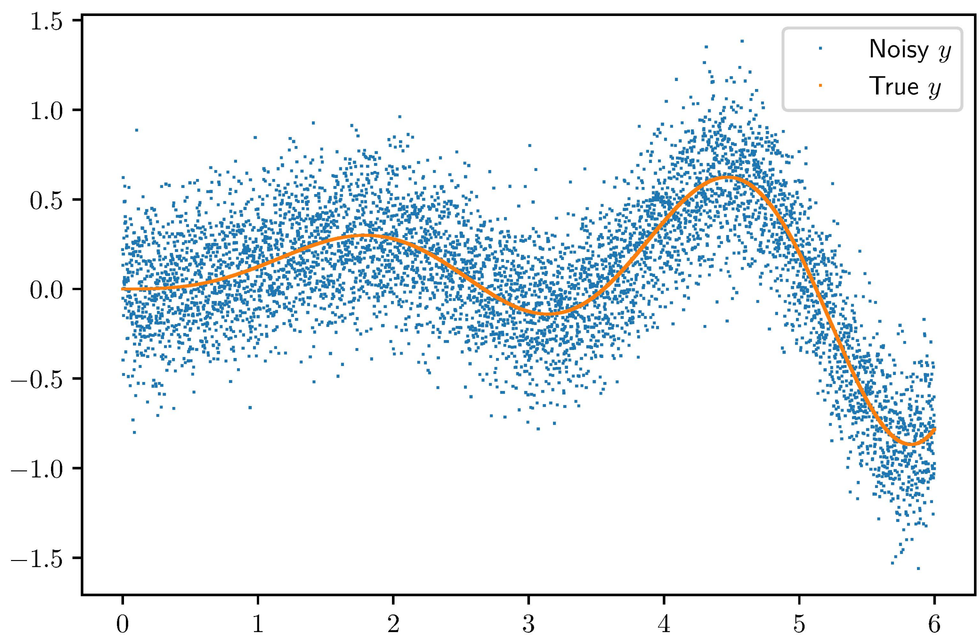

Figure 5.11 True and noisy system outputs with AWGN of $\mu = 0$ and $\sigma = 0.5$ and N=500

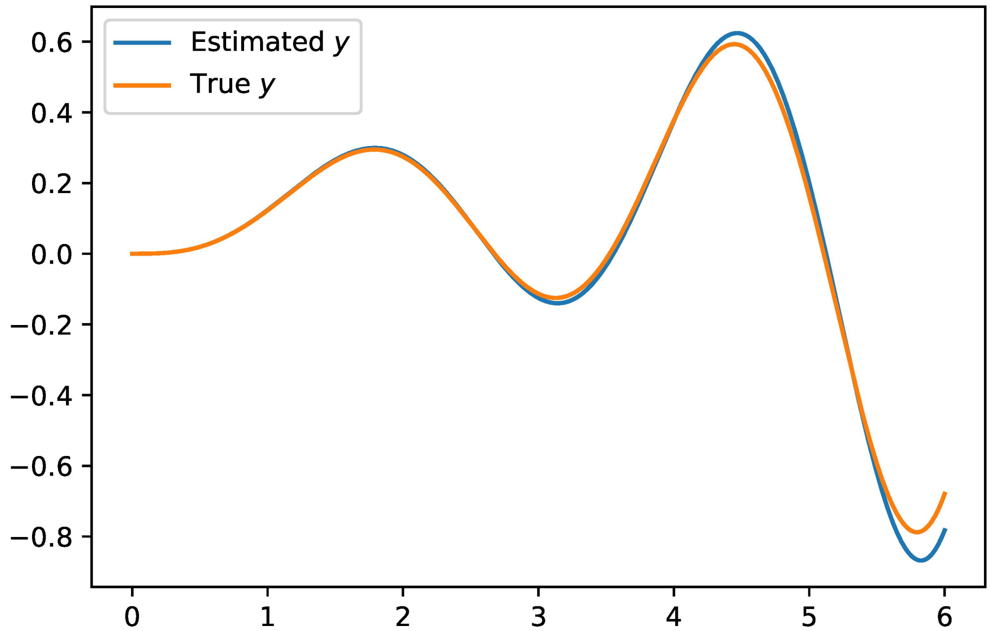

Figure 5.12 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 0.5$

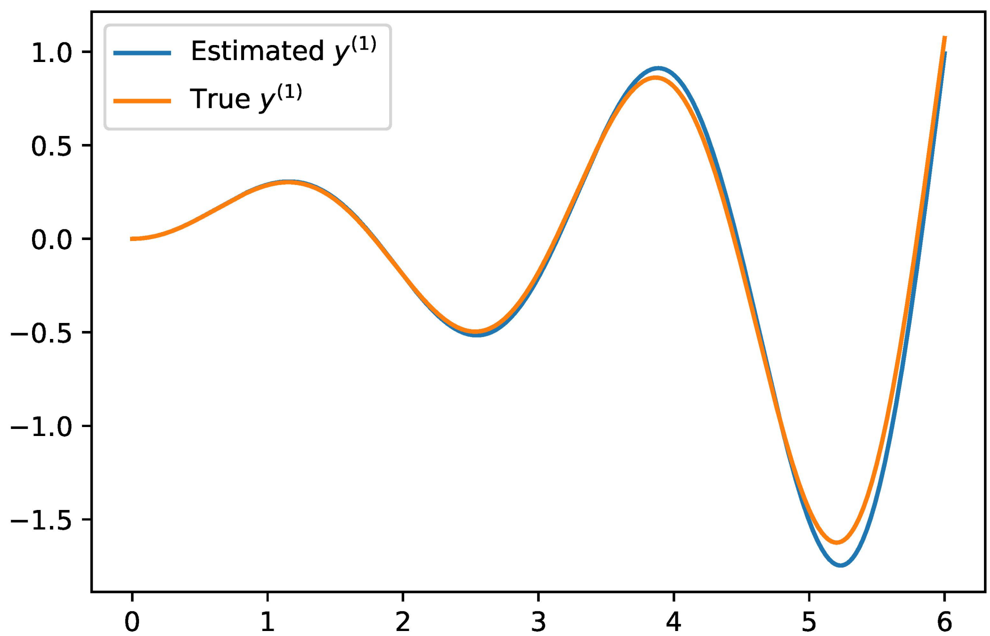

Figure 5.13 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.5$

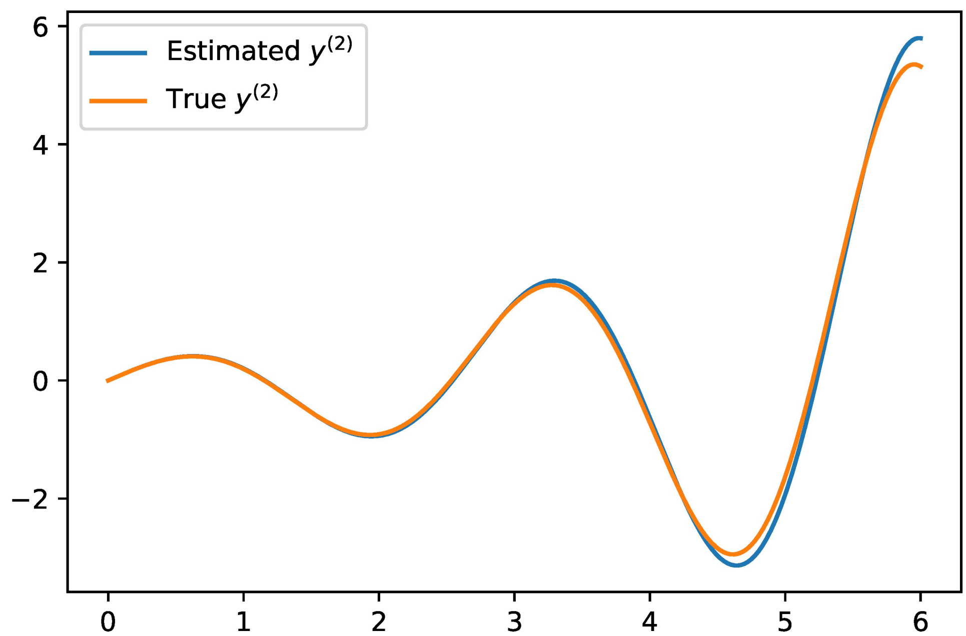

Figure 5.14 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.5$

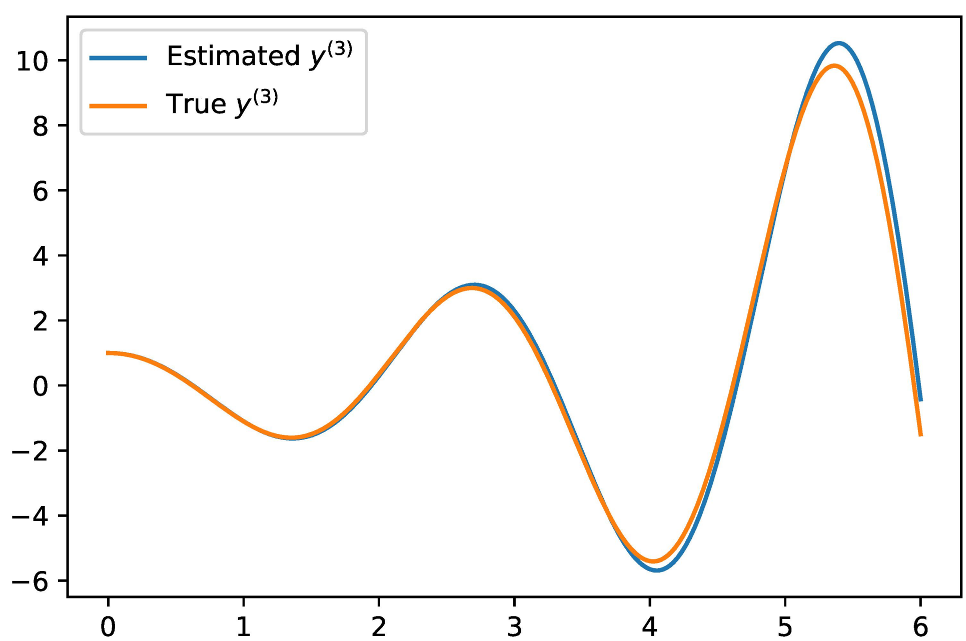

Figure 5.15 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.5$

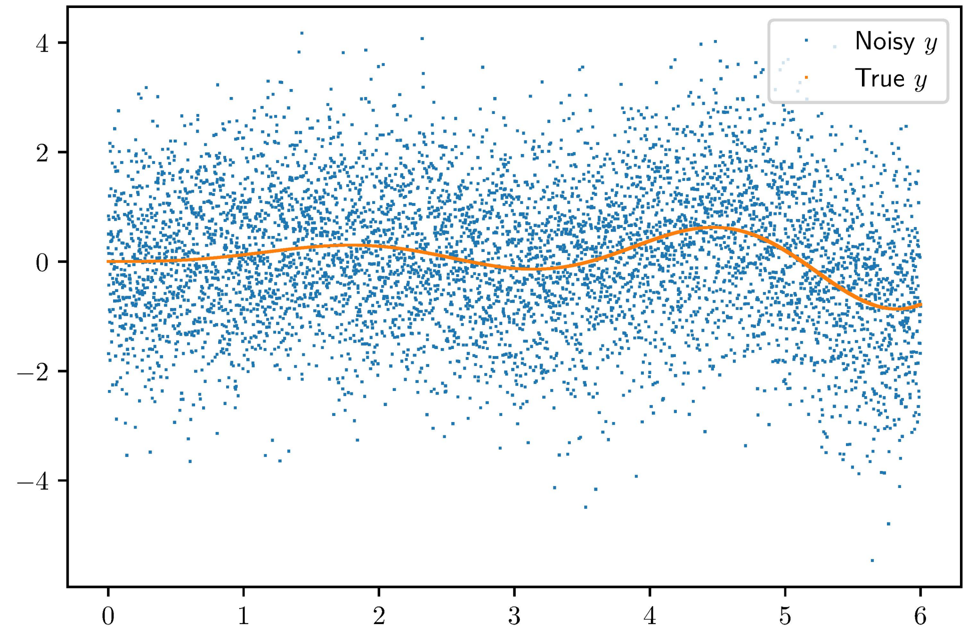

Figure 5.16 True and noisy system outputs with AWGN of $\mu = 0$ and $\sigma = 1.25$ and N=2000

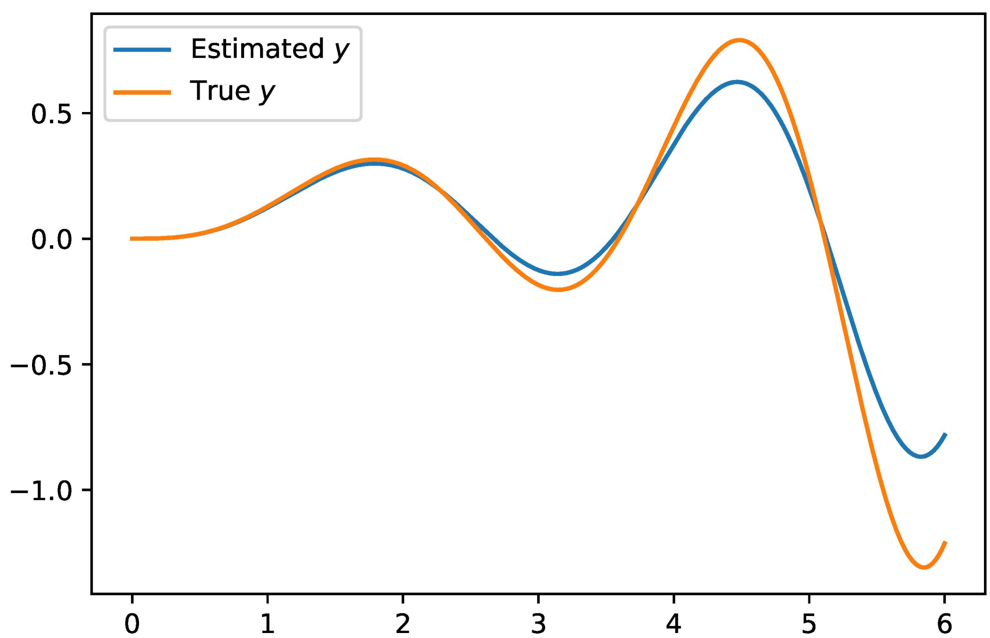

Figure 5.17 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 1.25$

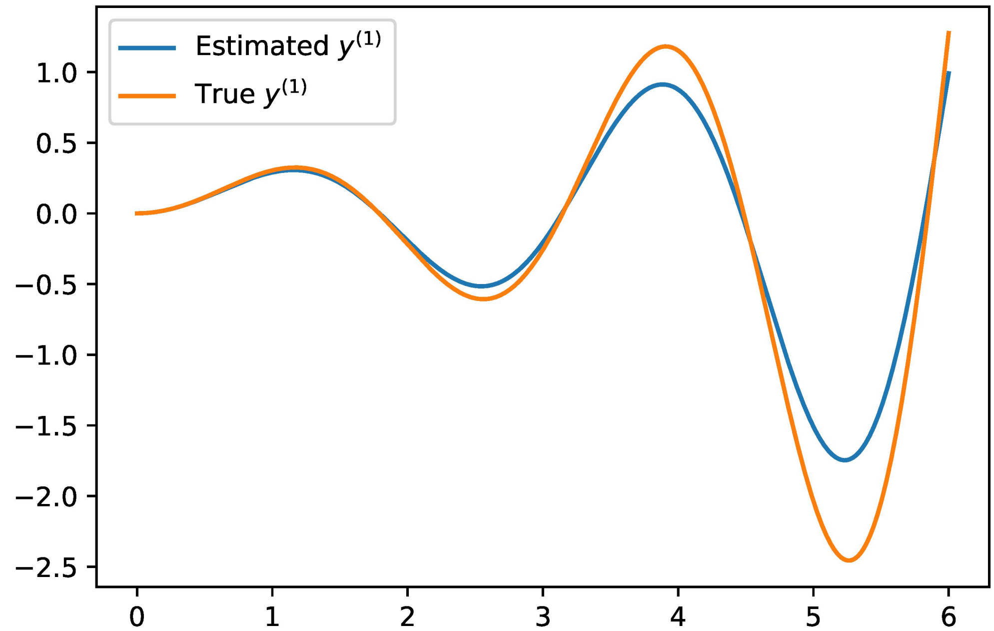

Figure 5.18 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.25$

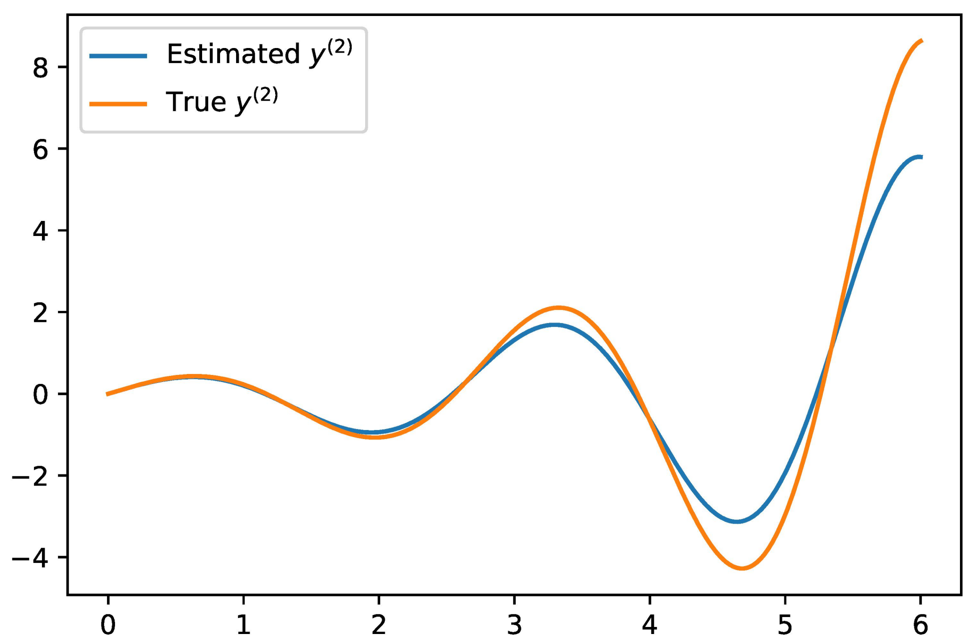

Figure 5.19 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.25$

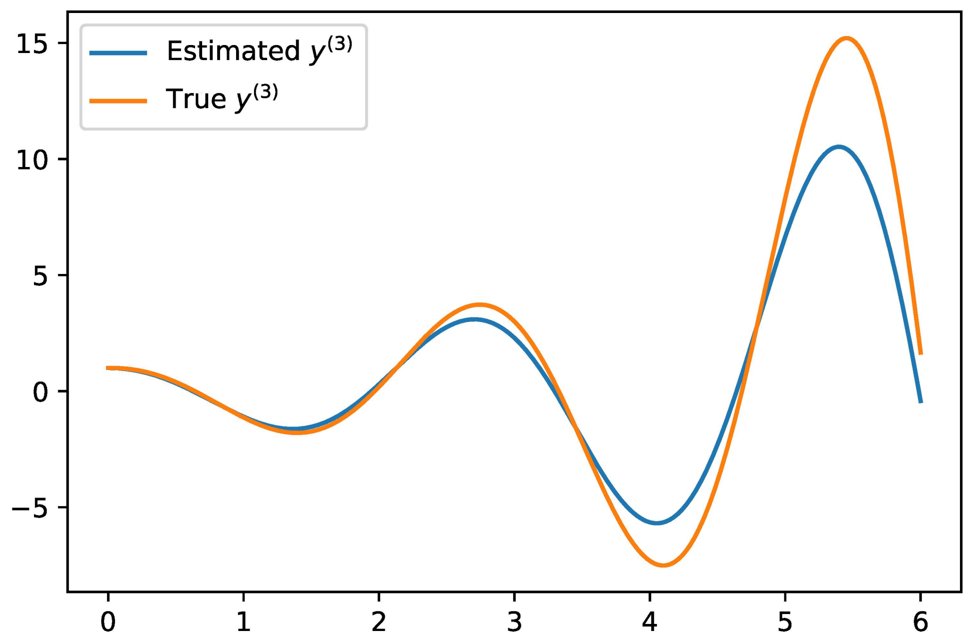

Figure 5.20 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.25$

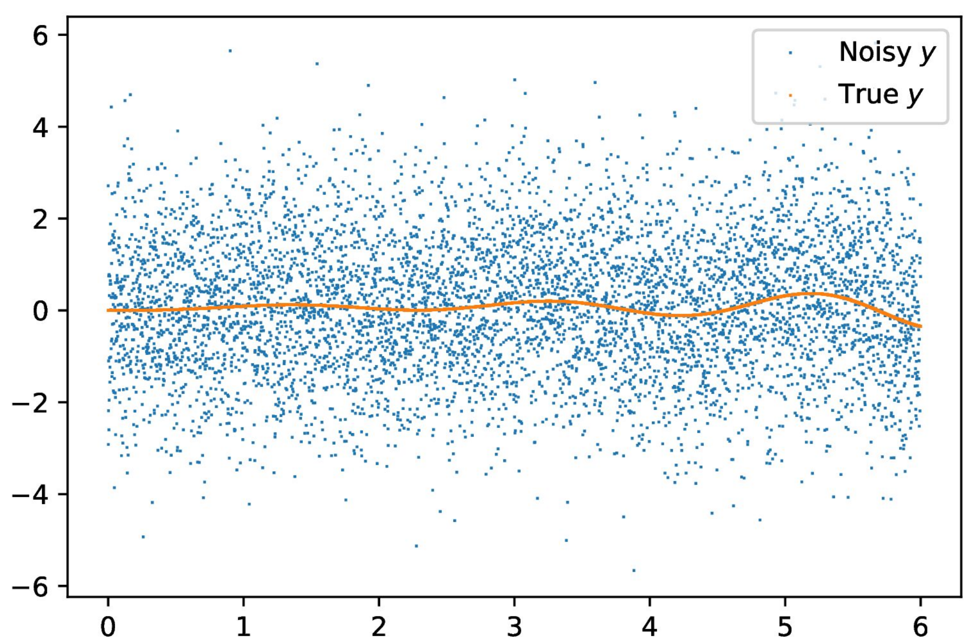

Figure 5.21 True and noisy system outputs with AWGN of $\mu = 0$ and $\sigma = 1.5$ and N=2000

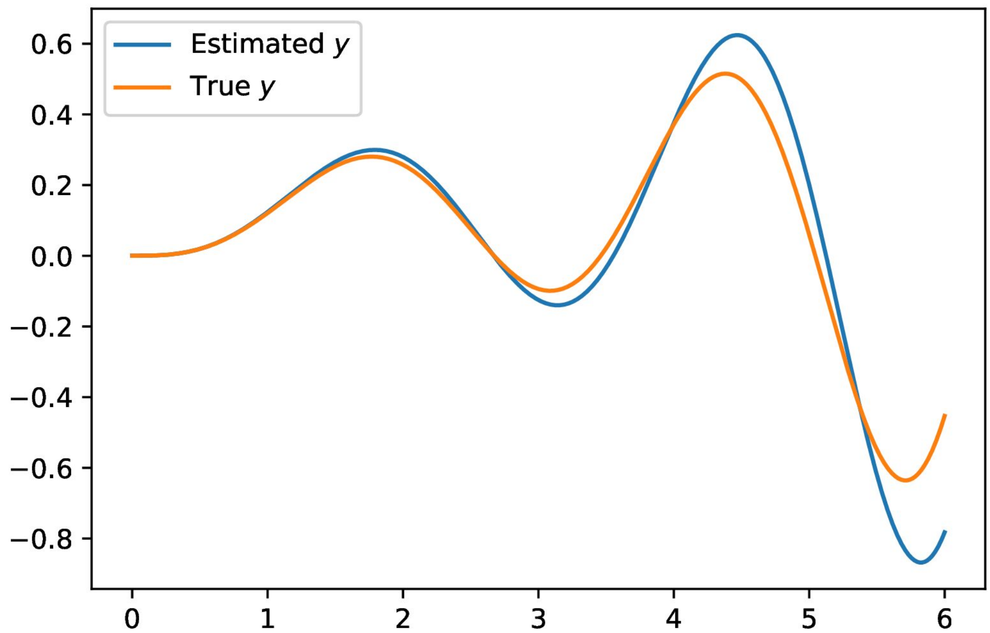

Figure 5.22 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 1.5$

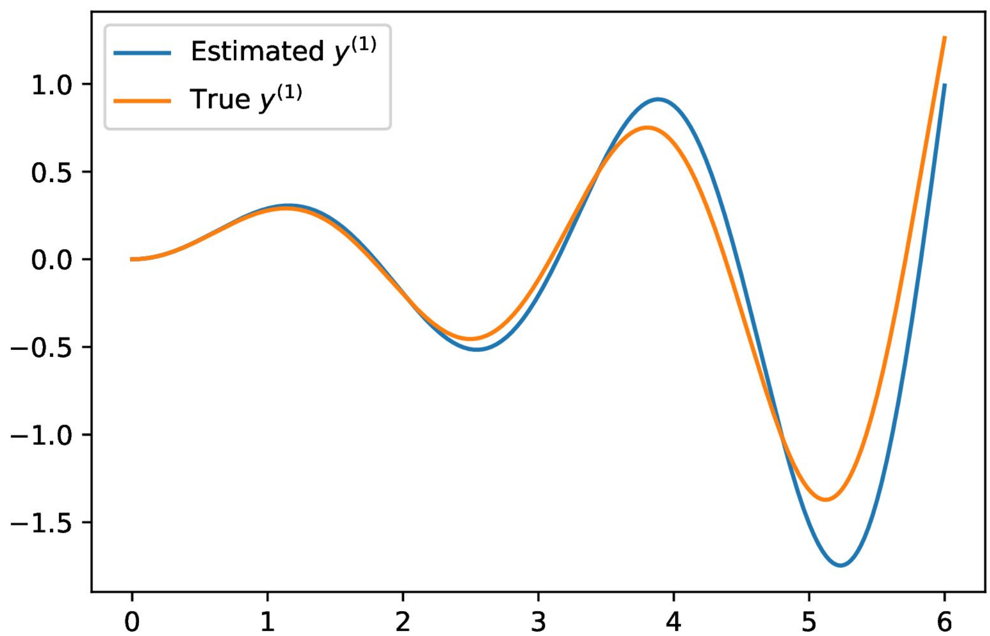

Figure 5.23 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.5$

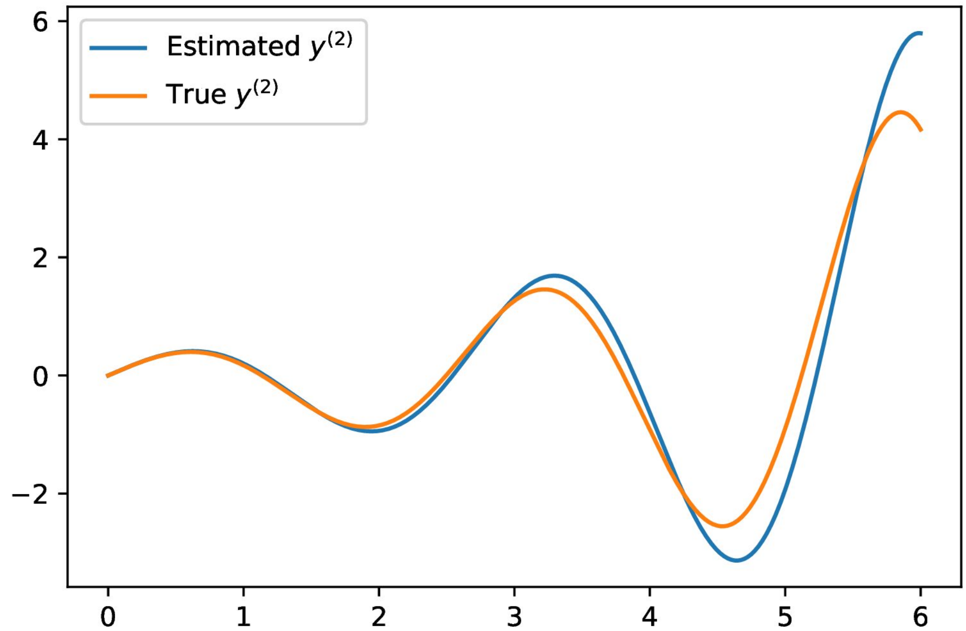

Figure 5.24 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.5$

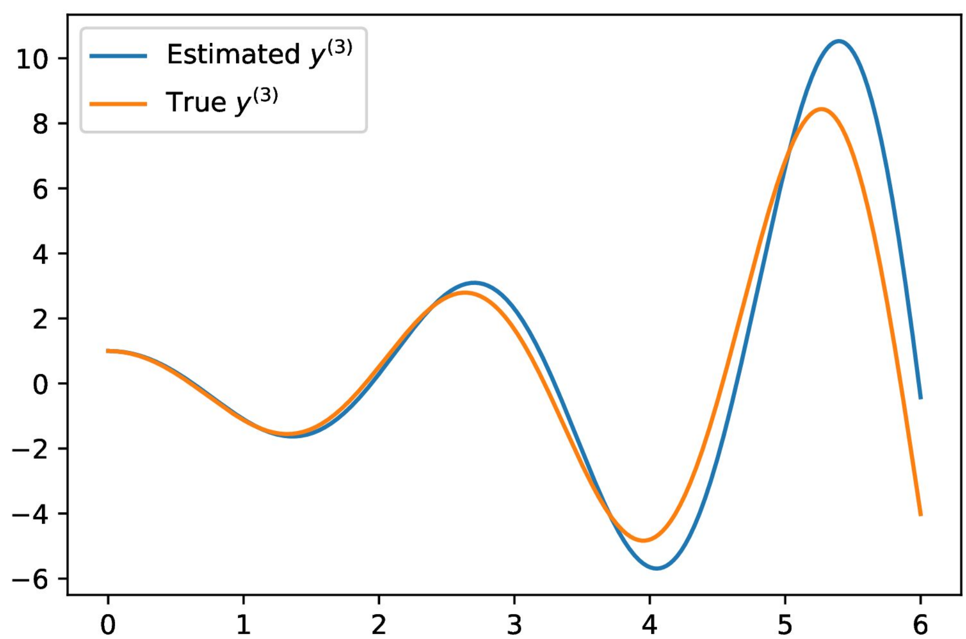

Figure 5.25 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1.5$

5.2 Discussion

5.2.1 Comparison with 'GLS-IV' [12] method

The FGLS algorithm adopted in this thesis is compared with the 'GLS-IV' described in [15]. The system considered is :

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -5 & -5 & -0 \end{bmatrix} x \;; y = x_1 \;;\; x(0) = [0,0,0,1] \tag{5.4}$$

with its corresponding characteristic equation

$$y^{(4)}(t) + a_3 y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{5.5}$$

where,

$$a_0 = 1,\ a_1 = 5,\ a_2 = 5,\ a_3 = 0$$

AWGN with variance 1 and 1.5(high noise) are added to the system. Below are the results obtained using both methods.

Variance	Knots(N)	Method	a_0	a_1	a_2	a_3
1	500	Ghosal et al. [12]	0.9464	4.8137	5.1652	0.1294
		Proposed method	0.98	4.96	4.910	-0.090
1.5	3000	Ghosal et al. [12]	0.5090	4.2273	5.2360	-0.1905
	2000	Proposed method	1.12	4.9762	5.329	0.03

Table 5.3 Comparative study with Ghoshal et al. [12]

The method adopted in this thesis proves to be more accurate as it can be observed clearly from the table that with less number of samples the parameters converge more accurately compare with [12].Also when the system is subjected to high noise with σ=1.5 the proposed method in this thesis gives better estimates with High noise and less samples comparatively.

5.2.2 Comparative approach of the influence of measurement noise levels and number of sample points on estimation

The Following example has been chosen to compare the results and progress of research with John et al [18]. The two-step asymptotic method used in John et al [18] rely on the sample theory i.e the system relies on large number of samples for better accuracy where as in the proposed method the accuracy is independent of the number of samples.
Example2: [18]
Consider a fourth order system as described below:

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -10 & -10 & 0 \end{bmatrix} x \; ; y = x_1 \; ; \; x(0) = [0, 0, 0, 1] \tag{5.6}$$

with its corresponding characteristic equation

$$y^{(4)}(t) + a_3 y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{5.7}$$

The poles of this system are $0.4677 \pm 3.2502i, -0.8228, -0.1127$. Hence the system is an unstable system. Unstable systems are known to be difficult to estimate.

The parameters a_0, a_1, a_2,a_3 are assumed to be unknown.
Consider the system in (5.6), on the same time interval $[0, 6]$. Three different noise levels are employed to emulate noise-perturbed output measurement in (5.6). Simultaneously, the number of sample points (N) is varied to study its influence on the accuracy of estimation. Additionally, to better understand how the number of sample points (N) affects the accuracy of estimation we compute the root mean square deviation, RMSD, as:

$$RMSD = \sqrt{\sum_{i=1}^{N} \frac{1}{n} [y_T(t_i) - \hat{y}(t_i)]^2} \tag{5.8}$$

where $y_T(t_i)$ and $\hat{y}(t_i)$ are: the true system output and the estimated output at time instant t_i, respectively.

Table (5.2.2) shows how the parameter estimation accuracy varies with increasing noise levels and number of sample points N.

Variance	Samples(N)	Method	$a0$	$a1$	$a2$	$a3$	$RMSD$
0.01	600	John et al. [18]	1.0130	6.6908	10.3540	-0.3164	0.0028
0.01	1800	John et al. [18]	1.0298	9.5368	10.0113	-0.3164	0.0009
0.01	600	Proposed method	0.9969	9.9925	10.069	0.004	0.00001
0.1	600	John et al. [18]	0.6849	13.1225	9.6459	0.2734	0.0044
0.1	6000	John et al. [18]	1.018	9.8892	10.4164	-0.1837	0.0012
0.1	600	Proposed method	1.012	10.140	9.801	0.002	0.0002
1	600	John et al. [18]	2.2352	-6.3921	10.0714	-0.2155	0.0192
1	15000	John et al. [18]	1.0657	8.0940	10.0482	-0.3455	0.0035
1	600	Proposed method	0.9969	9.9925	10.069	0.004	0.0014

Table 5.4 Estimates of parameter values and $RMSD$ for various noise levels and sample size N

The true system parameters are $a_0 = 1, a_1 = 10, a_2 = 10, a_3 = 0$.
In the two Step Asymptotic method proposed by John.et al [18].It can be observed from the result that as we increase the variance σ of the noise, a larger number of sample points is needed to achieve similar accuracy of estimation.

It should also be noted that the above Proposed method for parameter estimation in this thesis does not show any dependency on the samples to estimate the value of the parameters even for a unstable system with high noise variance .Irrespective of noise levels the parameters are estimated with highest accuracy compared to anju et al . [18]. On the pages that follow, we present the true output with AWGN of $\sigma = 0.01$ and N=600, $\sigma = 0.1$ and N=600, $\sigma = 1$ and N=600, and the reconstruction of the system output and its three derivatives for the best parameter estimates computed.

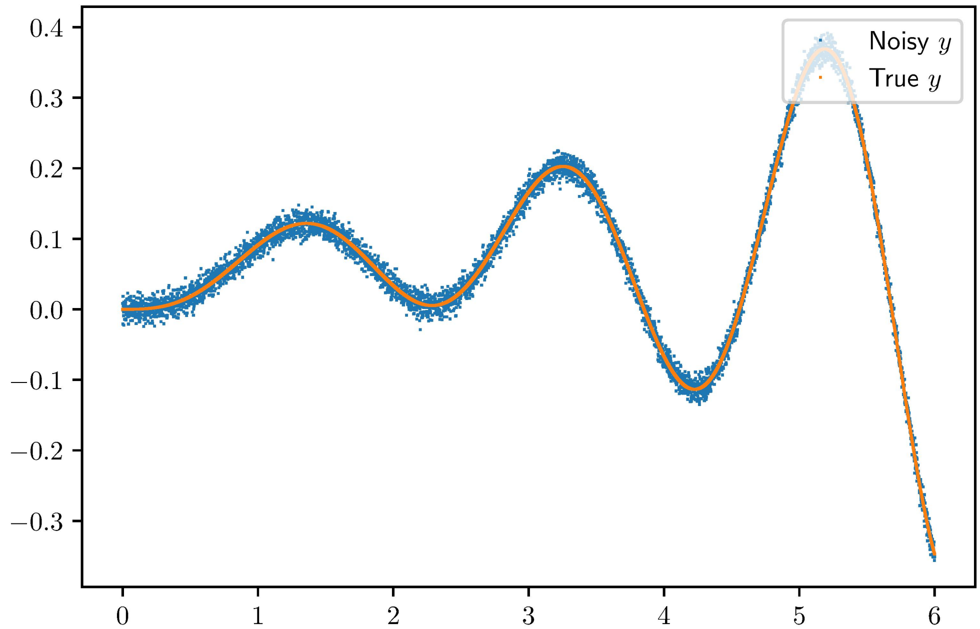

Figure 5.26 True and noisy system output with AWGN of $\mu = 0$ and $\sigma = 0.01$ and N=600

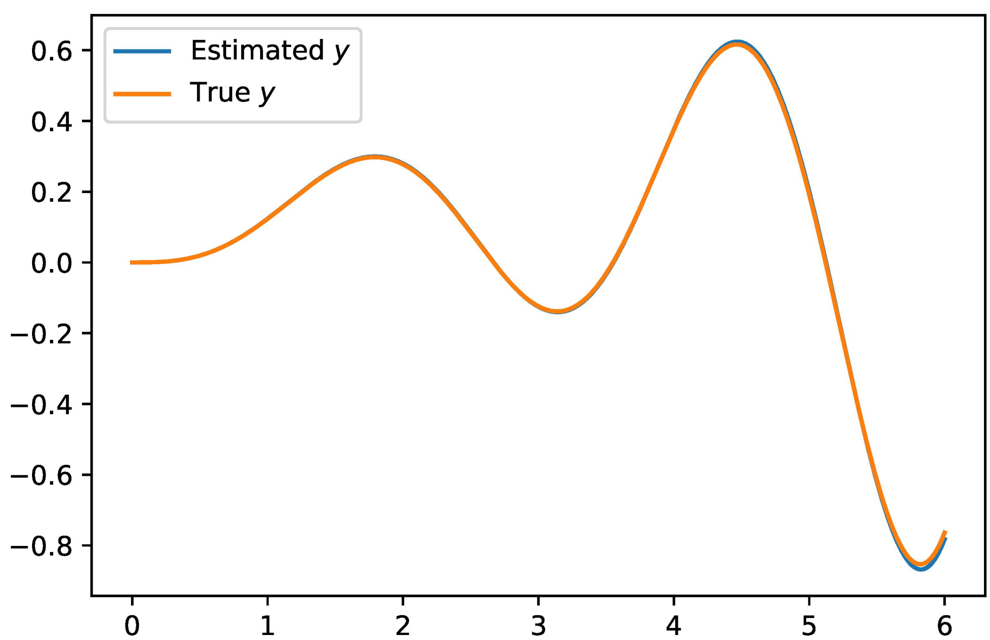

Figure 5.27 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 0.01$ and N=600

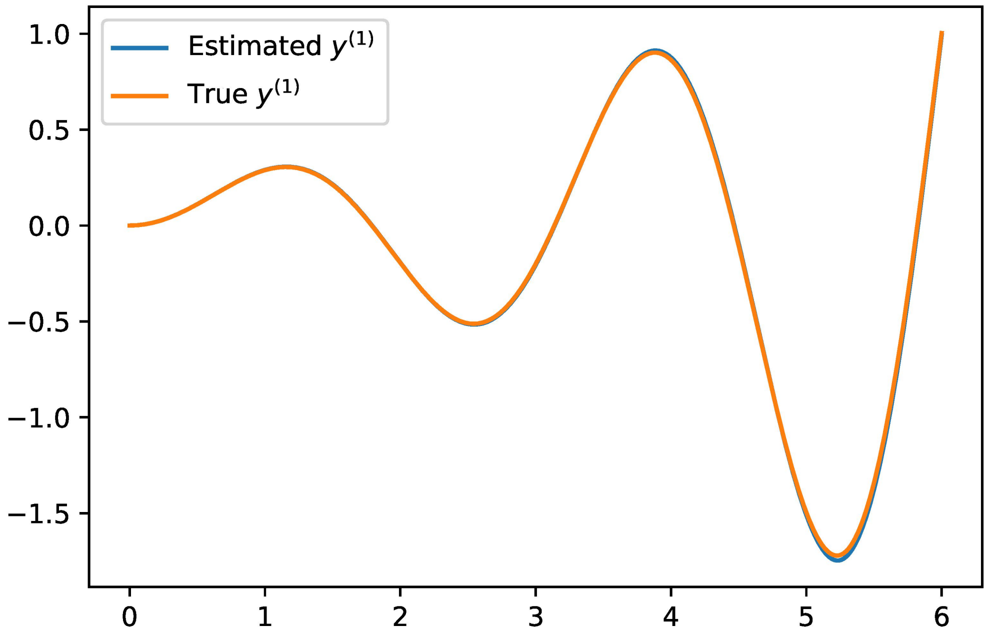

Figure 5.28 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.01$ and N=600

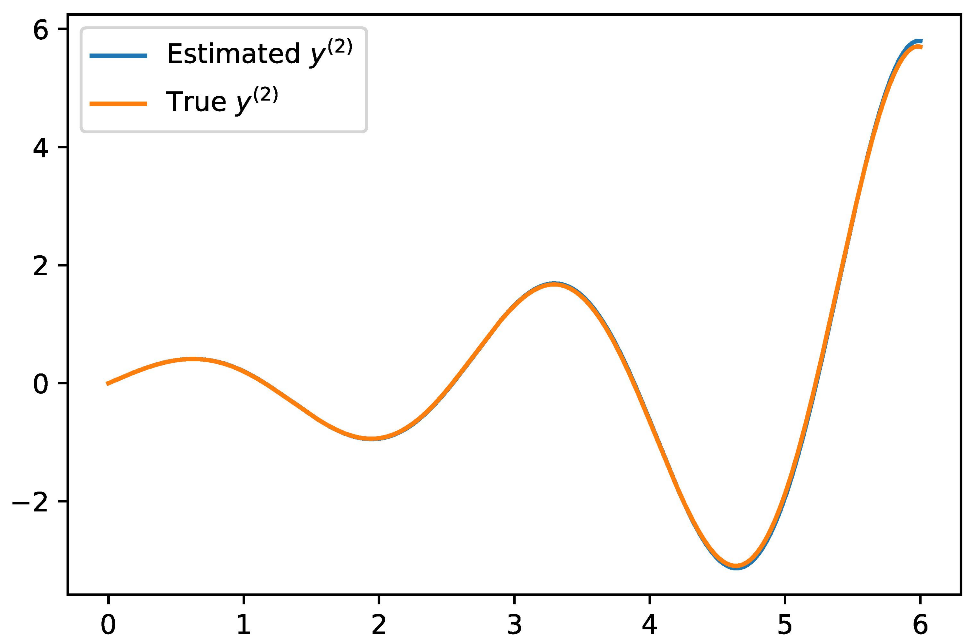

Figure 5.29 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.01$ and N=600

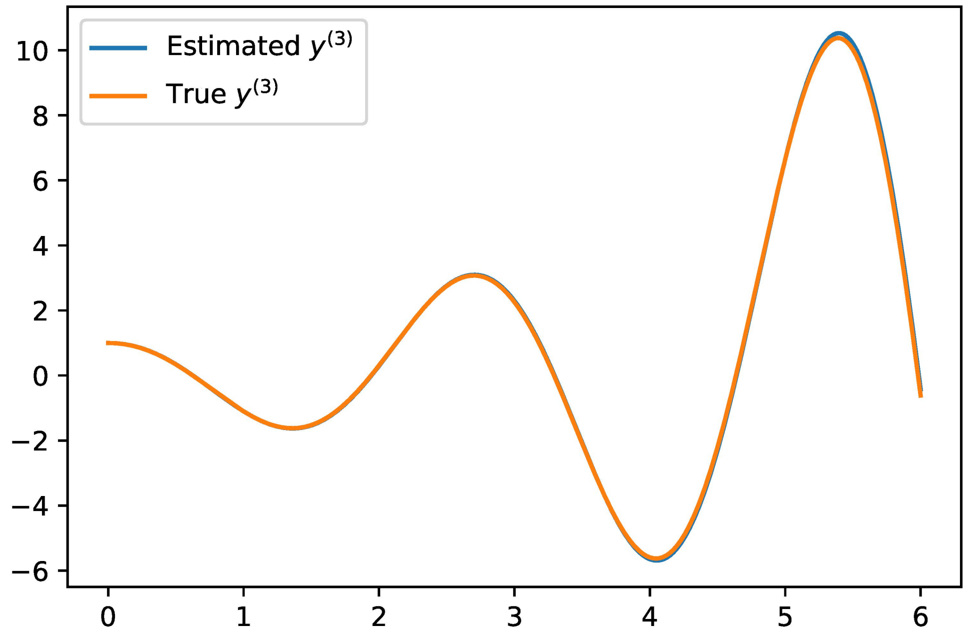

Figure 5.30 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.01$ and N=600

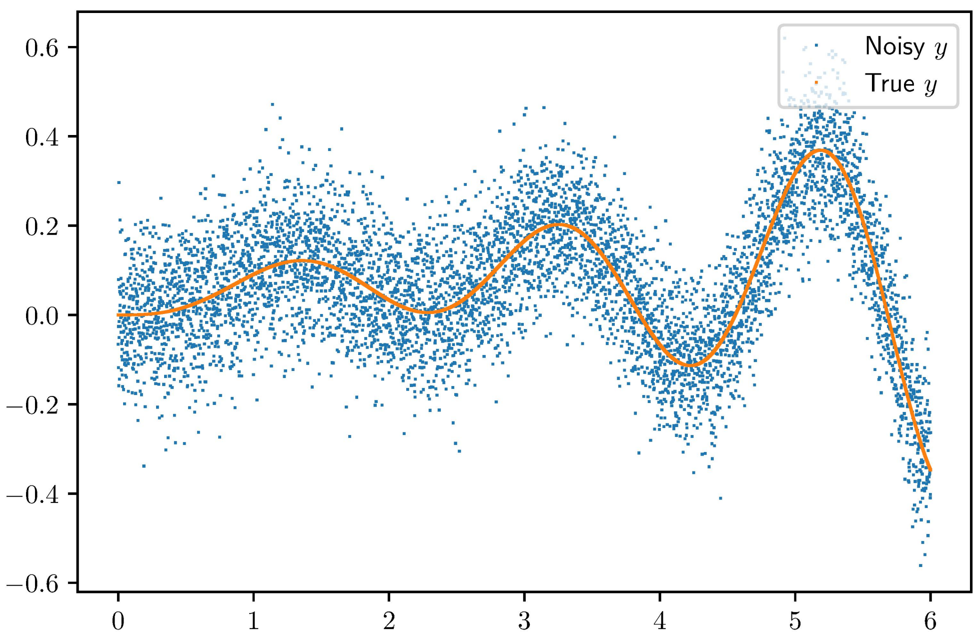

Figure 5.31 True and noisy system outputs with AWGN of $\mu = 0$ and $\sigma = 0.1$ and N=600

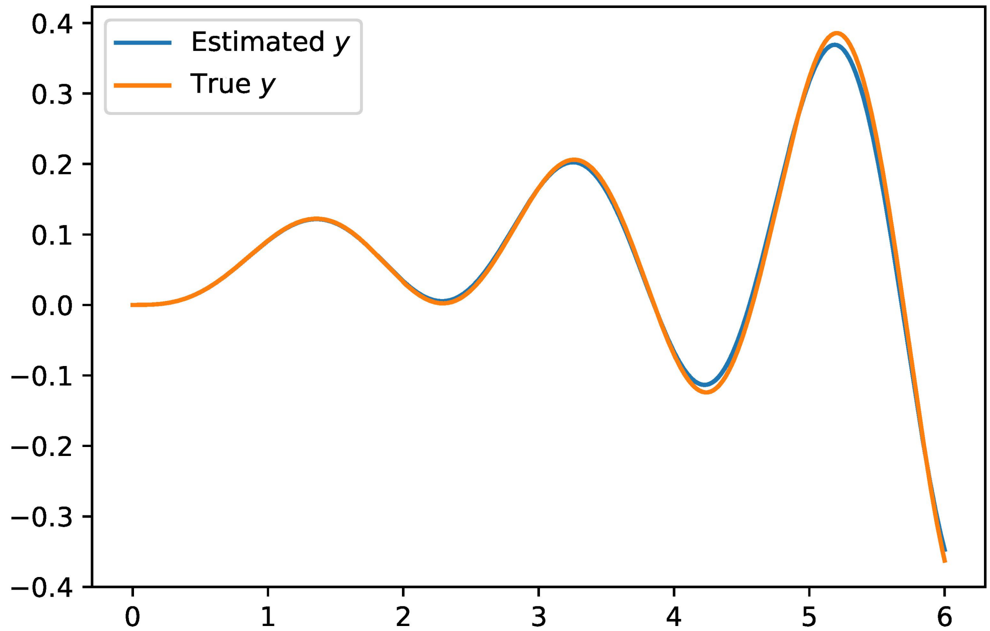

Figure 5.32 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 0.1$ and sample size N = 600

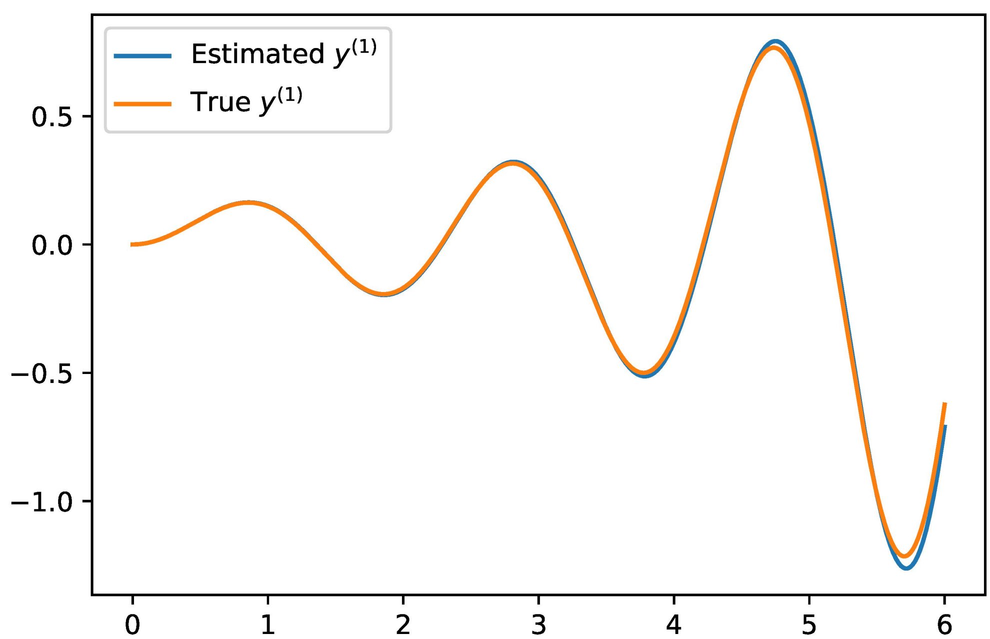

Figure 5.33 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1$ and N=600

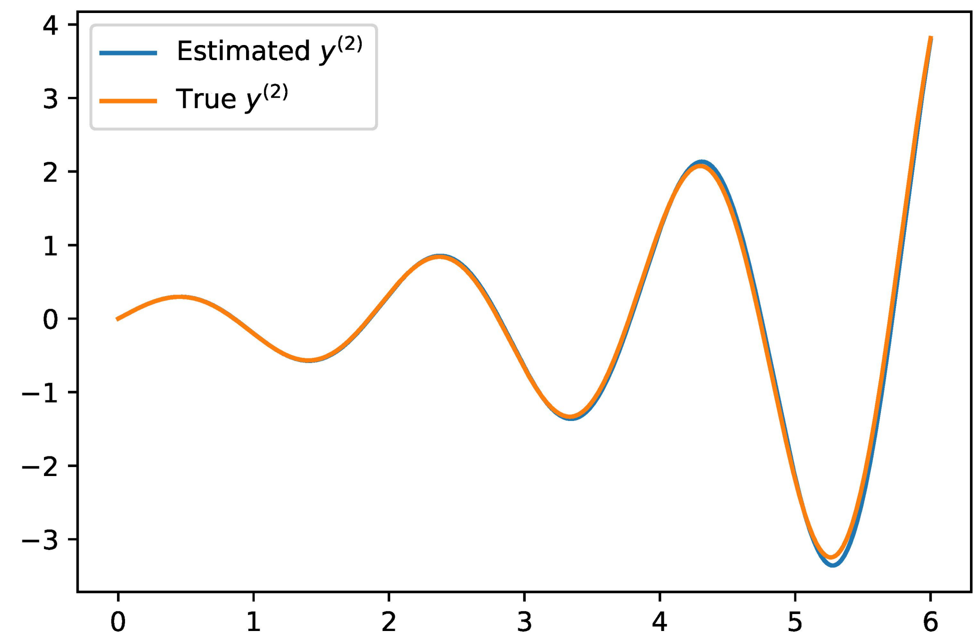

Figure 5.34 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.1$ and N=600

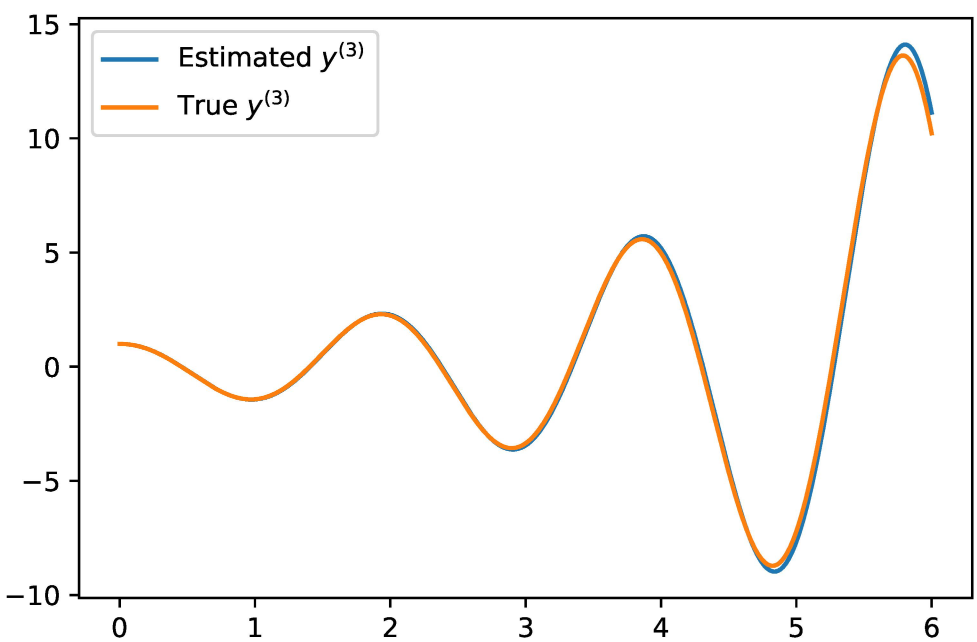

Figure 5.35 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 0.1$ and N=600

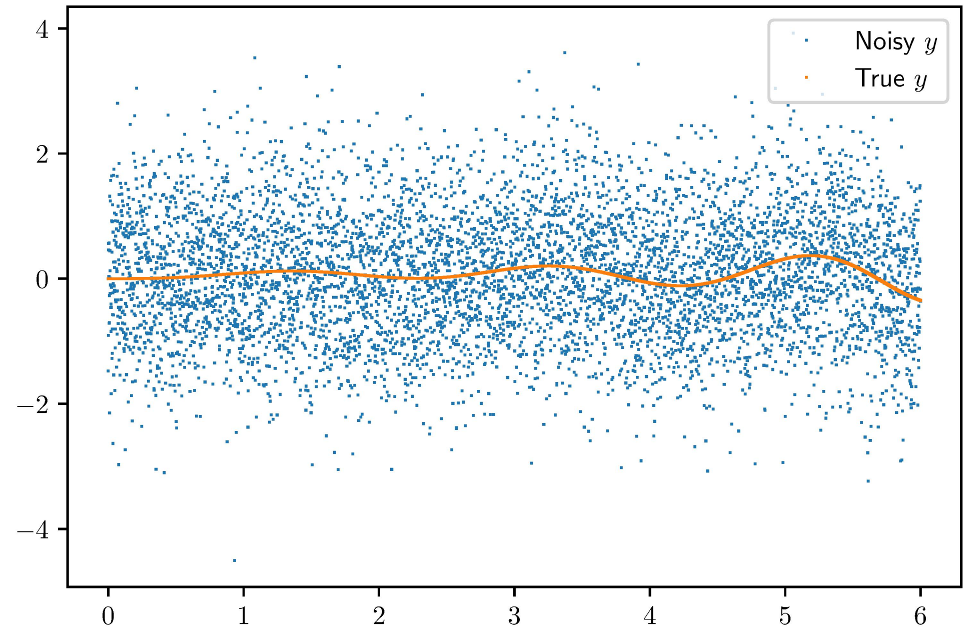

Figure 5.36 True and noisy system output with AWGN of $\mu = 0$ and $\sigma = 1$ and N=600

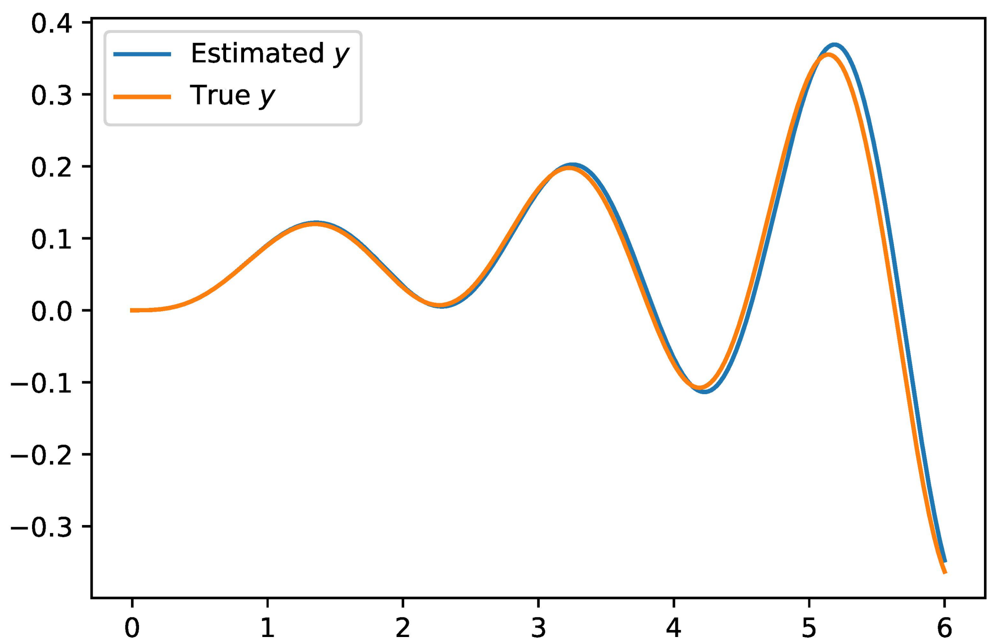

Figure 5.37 True and reconstructed output trajectories of the system with AWGN of $\mu = 0$ and $\sigma = 1$ and N=600

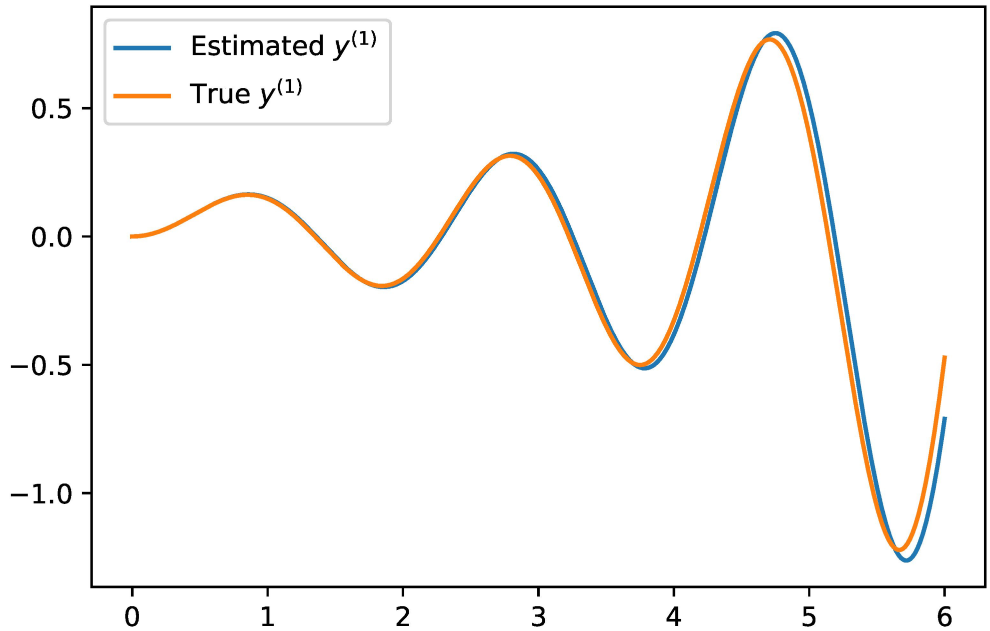

Figure 5.38 True and reconstructed first derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1$ and N=600

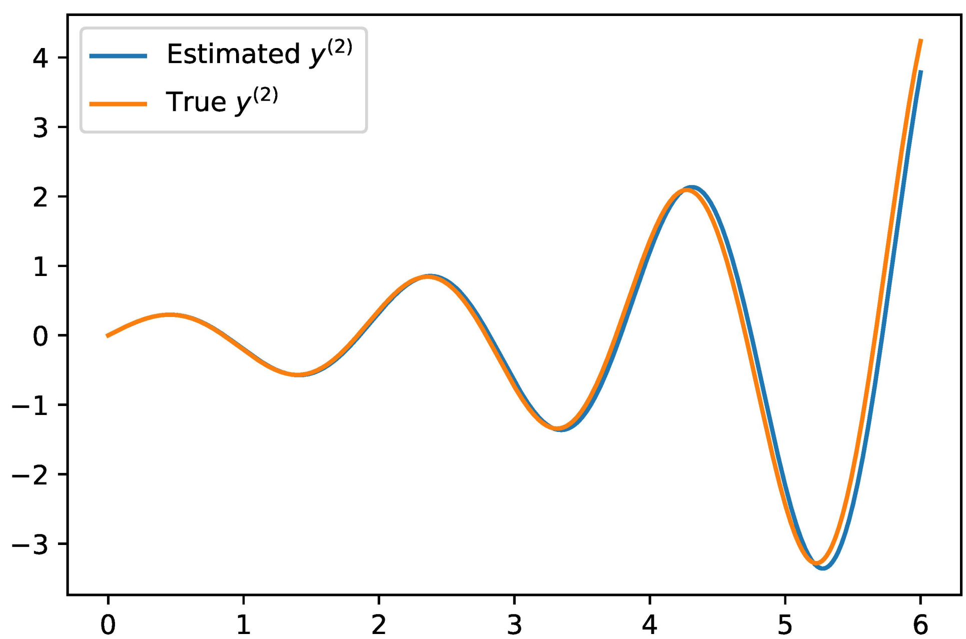

Figure 5.39 True and reconstructed second derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1$ and N=600

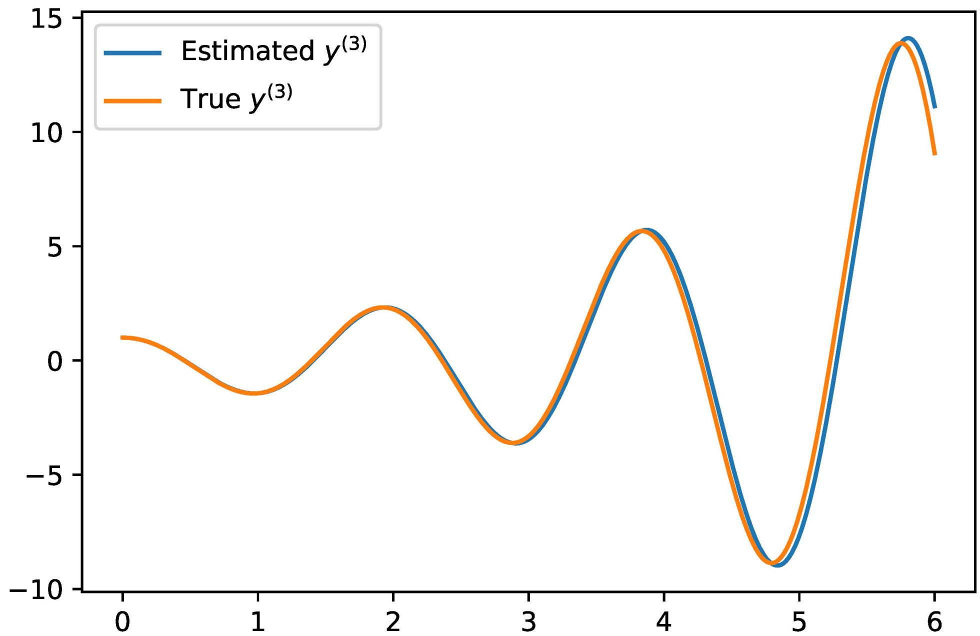

Figure 5.40 True and reconstructed third derivative of the system output with AWGN of $\mu = 0$ and $\sigma = 1$ and N=600

5.3 Summary

The state estimation, parameter estimation and filtering are extremely important to engineering applications. A comparison with previous works on kernel-based parameter and state estimation have been shown with results see section 5.2 and section 5.3 in this thesis to show the work progress and overcoming the limitations of previous works in our research group. To summarize the advantages of the proposed method:

*This method provides better parameter and state estimates with higher noise levels based on results shown in section 5.2.1

*This method does not rely on large sample theory as it is shown in section 5.2.2, it provides better results with lesser samples comparatively.

*It is not affected by initial conditions and hence the knowledge of initial conditions is not required.

Bibliography

[1] A. C. Aitken. Iv.—on least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1936.

KEY: aitken1936

[2] Siti Nora Basir, Hazlina Selamat, Hanafiah Yussof, Nur Ismarrubie Zahari, and Syamimi Shamsuddin. Parameter estimation of a closed loop coupled tank time varying system using recursive methods. *IOP Conference Series: Materials Science and Engineering*, 53:012052, dec 2013.

KEY: Basir2013

[3] P Caines. On the asymptotic normality of instrumental variable and least squares estimators. *IEEE Transactions on Automatic Control*, 21(4):598–600, 1976.

KEY: Caines1976

[4] Peter E Caines. *Linear stochastic systems*, volume 77. SIAM, 2018.

KEY: Caines2018

[5] Aritra Chatterjee. Two-stage kernel-based state and parameter estimation of lti systems. Master's thesis, McGill University, 2019.

KEY: Aritra

[6] Mohammed Dahleh, Munther A Dahleh, and George Verghese. Lectures on dynamic systems and control. $A+A$, 4(100):1–100, 2004.

KEY: Dahleh2004

[7] Michel Fliess and Hebertt Sira-Ramirez. Closed-loop parametric identification for continuous-time linear systems via new algebraic techniques. In H. Garnier & L. Wang, editor, *Identification of Continuous-time Models from Sampled Data*, Advances in Industrial Control, pages 362–391. Springer, 2008.

KEY: fliess:inria-00114958

[8] Michel Fliess and Hebertt Sira-Ramirez. Closed-loop parametric identification for continuous-time linear systems via new algebraic techniques. In *Identification of Continuous-time Models from sampled Data*, pages 363–391. Springer, 2008.

KEY: Fliess2008a

[9] Fliess, Michel and Sira-Ramírez, Hebertt. An algebraic framework for linear identification. *ESAIM: COCV*, 9:151–168, 2003.

KEY: fliess2003

[10] B Friedlander. The overdetermined recursive instrumental variable method. *IEEE Transactions on Automatic Control*, 29(4):353–356, 1984.

KEY: Friedlander1984

[11] Debarshi Ghoshal and Hannah Michalska. Finite-interval kernel-based identification and state estimation for lti systems with noisy output data. 07 2019.

KEY: Ghoshal2019

[12] Debarshi Patanjali Ghoshal and Hannah Michalska. Forward-backward kernel-based state and parameter estimation for linear systems of arbitrary order. *(submitted) International Journal of Control.*

KEY: Ghoshal

[13] DP Ghoshal, Kumar Gopalakrishnan, and Hannah Michalska. Algebraic parameter estimation using kernel representation of linear systems - submitted. In *IFAC World Congress, 2017. 20th World Congress*, 2017.

KEY: Ghoshal2017b

[14] DP Ghoshal, Kumar Gopalakrishnan, and Hannah Michalska. Using invariance to extract signal from noise. In *American Control Conference, 2017. ACC 2017*, 2017.

KEY: Ghoshal2017a

[15] DP Ghoshal and Hannah Michalska. Finite interval kernel-based identification and state estimation for lti systems with noisy output data. In *American Control Conference, 2019. ACC 2019*, 2019 forthcoming.

KEY: Ghoshal2017f

[16] Kumar Vishwanath Gopalakrishnan. A new method of non-asymptotic estimation for linear systems. Master's thesis, McGill University, 2016.

KEY: Kumar

[17] Simon Haykin. *Adaptive filter theory.* Prentice Hall, Upper Saddle River, NJ, 4th edition, 2002.

KEY: Haykin:2002

[18] Anju John. Estimation for siso lti systems using differential invariance. Master's thesis, McGill University, 2019.

KEY: john

[19] Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *Journal of basic engineering*, 83(1):95–108, 1961.

KEY: Kalman1961

[20] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(69):35–45, 1960.

KEY: Kalman1960

[21] D Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, 11.

KEY: Luenberger1966

[22] David G Luenberger. Observing the state of a linear system. *IEEE transactions on military electronics*, 8(2):74–80, 1964.

KEY: Luenberger1964

[23] DQ Mayne. A method for estimating discrete time transfer functions. In *Advances in Computer Control, Second UKAC Control Convention*. The University of Bristol, 1967.

KEY: Mayne1967

[24] Hannah Michalska. Instructor Notes: Parameter Estimation using Lsplines in RKHS, 2019.

KEY: Michalska1

[25] Abhishek Pandey. Deterministic variational approach to system modelling by data assimilation. Master's thesis, McGill University, 2018.

KEY: Pandey

[26] Hebertt Sira-Ramírez, Carlos García Rodríguez, John Cortés Romero, and Alberto Luviano Juárez. *Algebraic identification and estimation methods in feedback control systems*. John Wiley & Sons, 2014.

KEY: Ramirez2014b

[27] S Stanhope, JE Rubin, and David Swigon. Identifiability of linear and linear-in-parameters dynamical systems from a single trajectory. *SIAM Journal on Applied Dynamical Systems*, 13(4):1792–1815, 2014.

KEY: Stanhope2014

[28] Kwan Wong and Elijah Polak. Identification of linear discrete time systems using the instrumental variable method. *IEEE Transactions on Automatic Control*, 12(6):707–718, 1967.

KEY: Wong1967

[29] Peter C Young. The use of linear regression and related procedures for the identification of dynamic processes. In *Seventh Symposium on Adaptive Processes*, pages 53–53. IEEE, 1968.

KEY: Young1968

[30] Peter C Young. *Recursive estimation and time-series analysis: An introduction for the student and practitioner*. Springer Science & Business Media, 2011.

KEY: Young2011