TECHNISCHE UNIVERSITÄT
ILMENAU

# Optimal Input Design and Parameter Estimation for Continuous-Time Dynamical Systems

Dissertation

zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)

vorgelegt der

Fakultät für Informatik und Automatisierung

von

Simon Mayr
geboren am 16.04.1987 in Wels

| | |
|---|---|
| Gutachter: | Prof. Dr.-Ing. Johann Reger |
| | FH-Prof. DI Dr. Gernot Grabmair |
| | Prof. Dr. Jérôme Jouffroy |
| eingereicht am: | 02.07.2021 |
| verteidigt am: | 01.06.2022 |

**DOI:** 10.22032/dbt.52114
**URN:** urn:nbn:de:gbv:ilm1-2022000188

*Thank you to Prof. Dr.-Ing. Johann Reger and FH-Prof. DI Dr. Gernot Grabmair, for providing guidance, feedback, and patience throughout this project.*

*A big thank you to my wife, Kathrin, who has had my back during this time so I could focus on work.*

## Abstract

This thesis addresses two topics that play a significant role in modern control theory: design of experiments (DoE) and parameter estimation methods for continuous-time (CT) models. In this context, DoE focuses on the impact of experimental design regarding the accuracy of a subsequent estimation of unknown model parameters and applying the theory to real-world applications and its detailed analysis. We introduce the Fisher-information matrix (FIM), consisting of the parameter sensitivities and the resulting highly nonlinear optimization task. By a first-order system, we demonstrate the computation of the information content, its visualization, and an illustration of the effects of higher Fisher information on parameter estimation quality. After that, the topic optimal input design (OID), a subarea of DoE, will be thoroughly explored on the practice-relevant linear and nonlinear model of a 1D-position servo system. Comparison with standard excitation signals shows that the OID signals generally provide higher information content and lead to more accurate parameter estimates using least-squares methods. Besides, this approach allows taking into account constraints on input, output, and state variables.

In the second major topic of this thesis, we treat parameter estimation methods for CT systems, which provide several advantages to identify discrete-time (DT) systems, e.g., allows physical insight into model parameters. We focus on modulating function method (MFM) or Poisson moment functionals (PMF) and least-squares to estimate unknown model parameters. In the case of noisy measurement data, the problem of biased parameter estimation arises immediately. That is why we discuss the computation and compensation of the so-called estimation bias in detail. Besides the detailed elaboration of a bias compensating estimation method, this work's main contribution is, based on PMF and least squares for linear systems, the extension to at least slightly nonlinear systems. The derived bias-compensated ordinary least-squares (BC-OLS) approach for obtaining asymptotically unbiased parameter estimates is tested on a nonlinear 1D-servo model in the simulation and measurement. A comparison with other methods for bias compensation or avoidance, e.g., total least-squares (TLS), is performed. Additionally, the BC-OLS method is applied to the more general MFM. Furthermore, a practical issue of parameter estimation is discussed, which occurs when the system behavior leaves and re-enters the space covered by the identification equation. Using the 1D-servo system, one can show that disabling and re-enabling the PMF filters with appropriate initialization can solve this problem.

## Kurzfassung

Diese Arbeit behandelt die Themengebiete Design of Experiments (DoE) und Parameterschätzung für zeitkontinuierliche Systeme, welche in der modernen Regelungstheorie eine wichtige Rolle spielen. Im gewählten Kontext untersucht DoE die Auswirkungen von verschiedenen Rahmenbedingungen von Simulations- bzw. Messexperimenten auf die Qualität der Parameterschätzung, wobei der Fokus auf der Anwendung der Theorie auf praxisrelevante Problemstellungen liegt. Dafür wird die weithin bekannte Fisher-Matrix eingeführt und die resultierende nicht lineare Optimierungsaufgabe angeschrieben. An einem PT1-System wird der Informationsgehalt von Signalen und dessen Auswirkungen auf die Parameterschätzung gezeigt. Danach konzentriert sich die Arbeit auf ein Teilgebiet von DoE, nämlich Optimal Input Design (OID), und wird am Beispiel eines 1D-Positioniersystems im Detail untersucht. Ein Vergleich mit häufig verwendeten Anregungssignalen zeigt, dass generierte Anregungssignale (OID) oft einen höheren Informationsgehalt aufweisen und mit genaueren Schätzwerten einhergeht. Zusätzlicher Benefit ist, dass Beschränkungen an Eingangs-, Ausgangs- und Zustandsgrößen einfach in die Optimierungsaufgabe integriert werden können.

Der zweite Teil der Arbeit behandelt Methoden zur Parameterschätzung von zeitkontinuierlichen Modellen mit dem Fokus auf der Verwendung von Modulationsfunktionen (MF) bzw. Poisson-Moment Functionals (PMF) zur Vermeidung der zeitlichen Ableitungen und Least-Squares zur Lösung des resultierenden überbestimmten Gleichungssystems. Bei verrauschten Messsignalen ergibt sich daraus sofort die Problematik von nicht erwartungstreuen Schätzergebnissen (Bias). Aus diesem Grund werden Methoden zur Schätzung und Kompensation von Bias Termen diskutiert. Beitrag dieser Arbeit ist vor allem die detaillierte Aufarbeitung eines Ansatzes zur Biaskompensation bei Verwendung von PMF und Least-Squares für lineare Systeme und dessen Erweiterung auf (leicht) nicht lineare Systeme. Der vorgestellte Ansatz zur Biaskompensation (BC-OLS) wird am nicht linearen 1D-Servo in der Simulation und mit Messdaten validiert und in der Simulation mit anderen Methoden, z. B., Total-Least-Squares verglichen. Zusätzlich wird der Ansatz von PMF auf die weiter gefasste Systemklasse der Modulationsfunktionen (MF) erweitert. Des Weiteren wird ein praxisrelevantes Problem der Parameteridentifikation diskutiert, welches auftritt, wenn das Systemverhalten nicht gänzlich von der Identifikationsgleichung beschrieben wird. Am 1D-Servo wird gezeigt, dass ein Deaktivieren und Reaktivieren der PMF Filter mit geeigneter Initialisierung diese Problematik einfach löst.

# Contents

# List of Symbols and Acronyms

## Acronyms

**ACF** autocorrelation function

**AE** algebraic equation

**BC-OLS** bias-compensated ordinary least-squares

**BC-RLS** bias-compensated recursive least-squares

**BLUE** best linear unbiased estimator

**CCF** crosscorrelation function

**CDF** cumulative distribution function

**cond. PDF** conditional probability density function

**cond. PMF** conditional probability mass function

**CRLB** Cramér-Rao lower bound

**CRV** continuous random variable

**CT** continuous-time

**DoE** design of experiments

**DRV** discrete random variable

**DT** discrete-time

**ECU** electronic control unit

**EE** equation error

**EEM** equation error method

**ELS** extended least-squares

**FIM** Fisher-information matrix

**GLS** generalized least-squares

**i.i.d.** independent and identically distributed

**IIR** infinite impulse response

**IV** instrumental variables

**joint CDF** joint cumulative distribution function

**joint PDF** joint probability density function

**LTI** linear time-invariant

**MF** modulating function

**MFM** modulating function method

**MIMO** multiple-input multiple-output

**ML** Maximum Likelihood

**ODE** ordinary differential equation

**OE** output error

**OEM** output error method

**OID** optimal input design

**OLS** ordinary least-squares

**PCA** principal component analysis

**PDE** partial differential equation

**PDF** probability density function

**PEM** prediction error method

**PMF** Poisson moment functionals

**PMF** probability mass function

**PRBS** pseudorandom binary sequence

**PRMS** pseudorandom multilevel sequence

**RLS** recursive least-squares

**RP** random process

**RV** random variable

**SISO** single-input single-output

**SSS** strict sense stationary

**TLS** total least-squares

**WLS** weighted least-squares

**WSS** wide (weak) sense stationary

**ZOH** zero-order-hold

# Symbols

### Design of Experiments

| | |
|---|---|
| $\bar{\mathbf{F}}$ | average Fisher-information matrix |
| $\mathbf{\Lambda}_p$ | scaling matrix (diagonal matrix) |
| $\mathbf{F}^{-1}$ | Cramér-Rao Matrix (inverse Fisher-information matrix) |
| $\mathbf{F}$ | Fisher-information matrix |
| $\mathbf{J}$ | Jacobian matrix |
| $\mathbf{M}$ | parameter derivatives matrix |
| $\mathbf{S}$ | sensitivity matrix |
| $\tilde{\mathbf{F}}$ | scaled Fisher-information matrix |

### General

| | |
|---|---|
| $N$ | total number of samples |
| $T$ | sample time |
| $\mu$ | mean value |
| $\sigma$ | standard deviation |
| $s$ | complex s-domain variable (Laplace variable) |
| $z$ | complex z-domain variable |

### Mathematical and Stochastic Symbols & Operators

| | |
|---|---|
| $*$ | convolution operator |
| $\mathcal{L}^{-1}$ | inverse Laplace transform |
| $\mathcal{L}$ | Laplace transform |
| $\mathcal{N}$ | normally distributed |
| $\mathrm{E}\left[\cdot\right]$ | expected value (stochastic) |
| $\mathrm{cov}\left(\cdot\right)$ | covariance matrix |
| dlim | convergence in distribution |

| | |
|---|---|
| lim | limit in calculus (general) |
| plim | convergence in probability |
| $\text{var}(\cdot)$ | variance matrix |
| $f_{\text{x}}$ | probability density function (PDF) |

**Models**

| | |
|---|---|
| $\mathbf{A}$ | state (or system) matrix |
| $\mathbf{b}$ | model parameters, numerator coefficients of transfer function |
| $\mathbf{b}/\mathbf{B}$ | input vector/matrix |
| $\mathbf{c}^\top/\mathbf{C}$ | output vector/matrix |
| $\mathbf{f}(\cdot), \mathbf{h}(\cdot)$ | (sufficiently smooth) vector fields |
| $\mathbf{x}$ | state vector |
| $d/\mathbf{D}$ | feedthrough scalar/matrix |
| $m_d$ | $m_d + 1$ numerator coefficients (transfer-function) |
| $n_p$ | number of model parameters |
| $n_u$ | number of inputs |
| $n_y$ | number of outputs |
| $n$ | order of differential equation (model order); number of denominator coefficients (transfer-function) |
| $u/\mathbf{u}$ | model input (scalar/vector) |
| $y_{i,m,k}$ | $i$-th measured (noise corrupted) output, evaluated at $t_k = kT$ |
| $y/\mathbf{y}$ | model output (scalar/vector) |

**Parameter Estimation**

| | |
|---|---|
| $F^i(s)$ | continuous time filter performing $i$-th time derivative |
| $M_k\{\cdot\}$ | $k$-th order Poisson moment |
| $\Delta\mathbf{p}$ | estimation bias |
| $\mathbf{\Phi}$ | Matrix describing the contribution of filter coefficients with respect to the (asymptotic) estimation bias |

| | |
|---|---|
| $\hat{\mathbf{p}}$ | estimated parameter vector, $\hat{\mathbf{p}} = \begin{pmatrix}\hat{\mathbf{a}} & \hat{\mathbf{b}}\end{pmatrix}$ |
| $\mathbf{P}$ | parameter covariance matrix (least-squares algorithm) |
| $\mathbf{W}$ | data matrix or regressor matrix (least-squares algorithm) |
| $\mathbf{p}$ | nominal parameter vector, $\mathbf{p} = \begin{pmatrix}\mathbf{a} & \mathbf{b}\end{pmatrix}$ |
| $\mathbf{w}_k^\top$ | data vector or regressor vector (least-squares algorithm), $k-$th row of $\mathbf{W}$ |
| $\mathbf{y}_{LSQ}$ | observation vector (least-squares algorithm) |
| $\tilde{y}^{(i)}$ | $i$-th derivative of filtered output $y$ |
| $\tilde{y}$ | accent “tilde” indicates filtered signals, e.g., filtered output $y$ |
| $\varphi_k^{(i)}$ | $i$-th derivative of $k$-th order modulating function |
| $e$ | additive output noise |
| $g_F^i(t)$ | impulse response $g_F^i(t) = \mathcal{L}^{-1}\left(F^i(s)\right)$ |
| $n$ | white noise |
| $v/\mathbf{v}$ | (generalized) equation error |

# 1 Introduction

Methods and algorithms of modern control engineering and system theory are used in almost all technical areas, and their importance is even growing due to the increasing demands on intelligent systems. For example, this could be the automation or control of manufacturing and production processes and its online failure monitoring or high-speed positioning systems in the industrial sector. Another descriptive example from the automotive industry is the model-based estimation of vehicle dynamics or unknown parameters to improve vehicle control and stabilization or apply novel drive-train concepts, e.g., hybrid drive trains. Furthermore, intelligent mechatronic systems are also essential to overcome the arising challenges of the climate crisis. Control engineering methods and algorithms are used in all these systems, such as in smart grids with decentralized energy generators or failure monitoring in photovoltaic systems, to name but a few. The underlying mathematical models' quality significantly influences the performance of these control engineering methods and algorithms. For this reason, system identification is a must-have in control engineering. As stated in [1], the four essential tasks in system theory are modeling, analysis, estimation, and control. Of course, these tasks of control engineering interact with each other in many ways. From a system identification point of view, plant modeling and parameter estimation are essential, as these are directly related to the system identification problem. The concept of system identification describes the methodology of creating such mathematical models and consists of:

(i) obtaining input and output measurement data in time or frequency domain

(ii) selecting an appropriate model structure

(iii) estimating unknown model parameters

(iv) verifying the accuracy of the estimated model

Regarding choosing a suitable model structure (Item (ii)), there are two fundamentally different approaches: black-box and white-box models. They differ in the physical background used in the respective models. While white-box models are completely physics-based (theoretical modeling), black-box models are the opposite (experimental modeling), being entirely data-driven. Unsurprisingly, mixed forms are called gray-box models. Literature provides a good overview of the different types of models and their characteristics, e.g., [2,3]. In this thesis, the focus is not on the model structure's choice, and it is assumed that an appropriate model structure is obtained due to physics-based considerations. Measurement data generation (Item (i)) and the estimation of unknown model parameters (Item (iii)) remain the tasks to be solved.

This thesis addresses two selected issues in modern control theory; both of them inevitable for system identification. On the one hand, the generation of optimal experiments for dynamical systems, also known as DoE, and the estimation of unknown plant parameters, focusing on online capability, on the other hand. In terms of content, this work ranges from fundamental investigations on the topic of DoE for dynamical systems based on the FIM, initially introduced by [4], in Chapter 2, to methods for parameter estimation like modulating function (MF) approaches or PMF [5], calculation and compensation of estimation bias [5–7] in Chapter 3 and application of the investigated methods to selected practical applications in Chapter 4. Finally, the necessary fundamentals of random variables (RVs) and random processes (RPs) are treated in Section A.1 and Section A.3.

At first glance, DoE and parameter estimation are thematically decoupled research topics, which is why they are dealt with separately in Chapter 2 and Chapter 3, respectively. However, it appears obvious to use DoE methods to improve parameter estimation results, independent of the chosen approach later on for parameter estimation. DoE includes, among other things, the generation of input signals in compliance with restrictions, e.g., mechanical, the finding of optimal measurement time instances, or the optimal placement of sensors to obtain more accurate parameter estimates. DoE is often used to generate excitation (input) signals or informative input signals adjusted to the respective system. We call this subdomain optimal input design (OID).

As a first introduction to the topic of DoE, we illustrate the effect of an arbitrary choice of evaluation time instances on parameter estimates of a simple first-order system. We consider a first-order single-input single-output (SISO) linear time-invariant (LTI) system with transfer function

$$G(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \frac{V}{1 + sT_1} \tag{1.1}$$

where $s$ describes the Laplace variable. Its physics-based model parameters $V$ (gain) and $T_1$ (time constant) should be estimated as best possible. The nominal model parameters are $V = 3$ and $T_1 = 1$. The system is assumed to be excited with a unit step $u(t) = \sigma(t)$ for which the system response in time-domain $y(t)$ is called step-response, and can be analytically derived by, e.g., application of the inverse Laplace transform

$$\begin{aligned} y(t) &= \mathcal{L}^{-1}\left(\hat{u}(s)\, G(s)\right) \\ &= \mathcal{L}^{-1}\left(\frac{1}{s}\frac{V}{1+sT_1}\right) = V\left(1 - e^{-\frac{t}{T_1}}\right) . \end{aligned} \tag{1.2}$$

Furthermore, we assume that the measured output $y_m(t)$ is sampled at discrete time instances $t_k = kT$, $1 \leq k \leq N$, with sample time $T$ and the total number of samples $N$, and noise corrupted with $e_k \sim \mathcal{N}(0, \sigma^2)$, where $\sigma = 0.1$, i.e.,

$$y_m(t_k) = y_{m;k} = y_k + e_k . \tag{1.3}$$

The input and output signals are visualized in Fig. 1.1. To determine the unknown

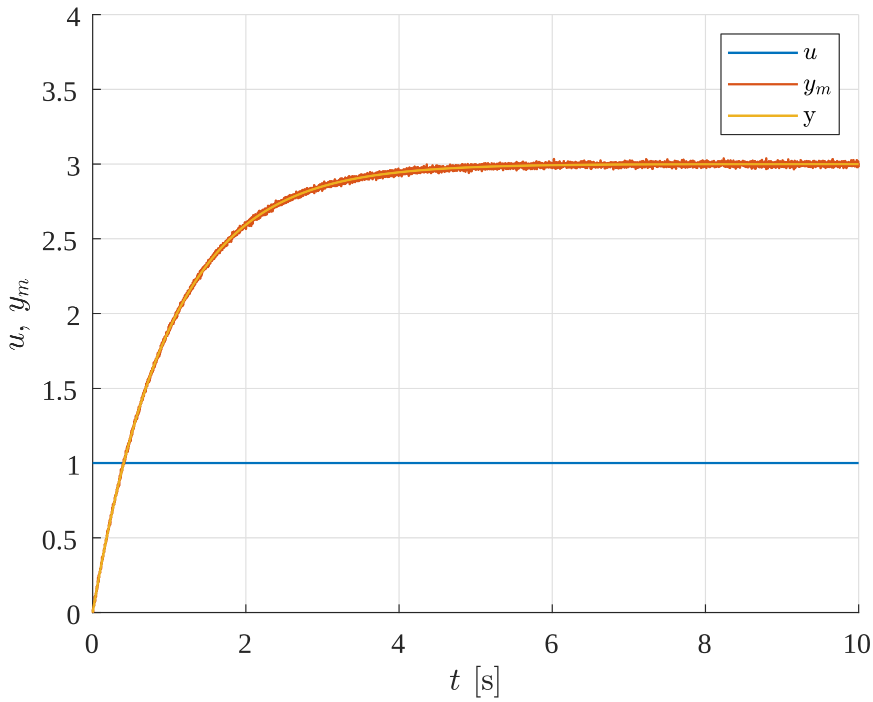



Figure 1.1: Step response $y_m(t) = y(t) + e(t)$, where $e \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$ .

parameters, (1.2) must be evaluated at two different $t_1$ and $t_2$, respectively. When looking at (1.2) or Fig. 1.1, it is evident that the DC-gain $V$ can be estimated by tending $t_2$ towards infinity

$$\hat{V} = \lim_{t_2 \to \infty} V\left(1 - e^{-\frac{t_2}{T_1}}\right) = y_m\left(t_2\right) = y_{m;2} \tag{1.4}$$

with $t_2 > t_1$. For the time constant, one obtains

$$\hat{T}_1 = -\frac{t_1}{\ln\left(1 - \frac{y_{m;1}}{\hat{V}}\right)} = -\frac{t_1}{\ln\left(1 - \frac{y_{m;1}}{y_{m;2}}\right)} . \tag{1.5}$$

The choice of a suitable, or even optimal, time instance $t_1$ remains open. To show the effects of different evaluation time pairs $\{t_1, t_2\}$ on the estimation result $\hat{V}$, $\hat{T}_1$, and to demonstrate the effectiveness of DoE methods, three different value pairs and the respective resulting parameter estimates are compared in the following. We take the liberty to anticipate Chapter 2 and briefly introduce the DoE method for dynamical systems without presenting detailed background knowledge and mathematical derivations. Of course, this is discussed in much more detail in Chapter 2. In a nutshell, the DoE method for dynamical systems generates experiments that are as informative as possible. This information content is described by the so-called Fisher-information matrix (FIM) or a scalar quantity of the same matrix. The FIM is mainly composed of system outputs' sensitivities concerning selected, maybe unknown, model parameters. Thus, the FIM applied to (1.2) with nominal parameter vector $\mathbf{p}_0 = (V, T_1)^\top$ read as

$$\begin{aligned}
\mathbf{F}(u, \mathbf{p}_0)\big|_{u=\sigma(t)} &= \sum_{k=1}^{2} \frac{1}{\sigma^2} \begin{pmatrix} \left(\frac{\partial y}{\partial V}\right)^2\Big|_{\mathbf{p}_0, t_k} & \left(\frac{\partial y}{\partial V}\frac{\partial y}{\partial T_1}\right)\Big|_{\mathbf{p}_0, t_k} \\ \left(\frac{\partial y}{\partial T_1}\frac{\partial y}{\partial V}\right)\Big|_{\mathbf{p}_0, t_k} & \left(\frac{\partial y}{\partial T_1}\right)^2\Big|_{\mathbf{p}_0, t_k} \end{pmatrix} \\
&= \frac{1}{\sigma^2} \sum_{k=1}^{2} \begin{pmatrix} \left(1 - e^{-t_k/T_1}\right)^2 & \left(1 - e^{-t_k/T_1}\right)\left(-\frac{t_k V}{T_1^2} e^{-\frac{t_k}{T_1}}\right) \\ \left(1 - e^{-t_k/T_1}\right)\left(-\frac{t_k V}{T_1^2} e^{-\frac{t_k}{T_1}}\right) & \left(-\frac{t_k V}{T_1^2} e^{-\frac{t_k}{T_1}}\right)^2 \end{pmatrix}
\end{aligned} \tag{1.6}$$

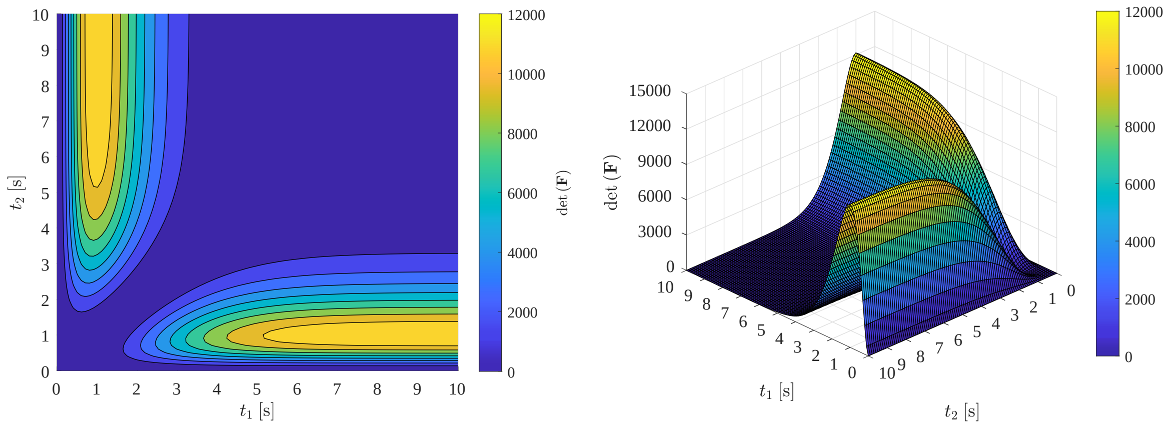
with the respective local sensitivities

$$\begin{aligned}
s_1\left(t_k\right) &= \frac{\partial y}{\partial V}\bigg|_{\mathbf{p}_0, t_k} = 1 - e^{-\frac{t_k}{T_1}} \\
s_2\left(t_k\right) &= \frac{\partial y}{\partial T_1}\bigg|_{\mathbf{p}_0, t_k} = -\frac{t_k V}{T_1^2} e^{-\frac{t_k}{T_1}}
\end{aligned} \tag{1.7}$$

and output noise variance $\sigma^2 = 0.01$ . For optimization purposes, the determinant of the FIM is used as a scalar objective function. We end up with the constrained optimization problem

$$\max_{t_1,t_2} \det\left(\mathbf{F}\left(u, \mathbf{p}_0\right)\right)$$
$$t_1 \in [0, 10] \tag{1.8}$$
$$t_2 \in [0, 10]$$

to obtain optimal time instances.[1] In Fig. 1.2, the obtained information content (1.8) is shown as a function of the chosen time instances $t_1$ and $t_2$. It is easy to see that the information content reaches a maximum with $\{t_1, t_2\} = \{1\,\text{s}, 10\,\text{s}\}$ or $\{t_1, t_2\} = \{10\,\text{s}, 1\,\text{s}\}$, respectively. The prior knowledge to tend $t_2 \to \infty$, as assumed in (1.4), is also a direct result of the DoE task. The DoE result seems quite plausible since it




(a) Contour plot.

(b) Three-dimensional surface plot.

Figure 1.2: Optimal measurement points $t_1$ and $t_2$ for estimating the model parameters $V$ and $T_1$ using D-criterion for evaluating the Fisher matrix.

means that the gain factor $V$ can be determined best in the steady-state range and the time constant $T_1$ within the range of large slope.

To determine whether the parameter estimation (1.4) and (1.5) is affected by choice of evaluation time instances thousand simulation runs are performed, each evaluated at three different pair of time values:

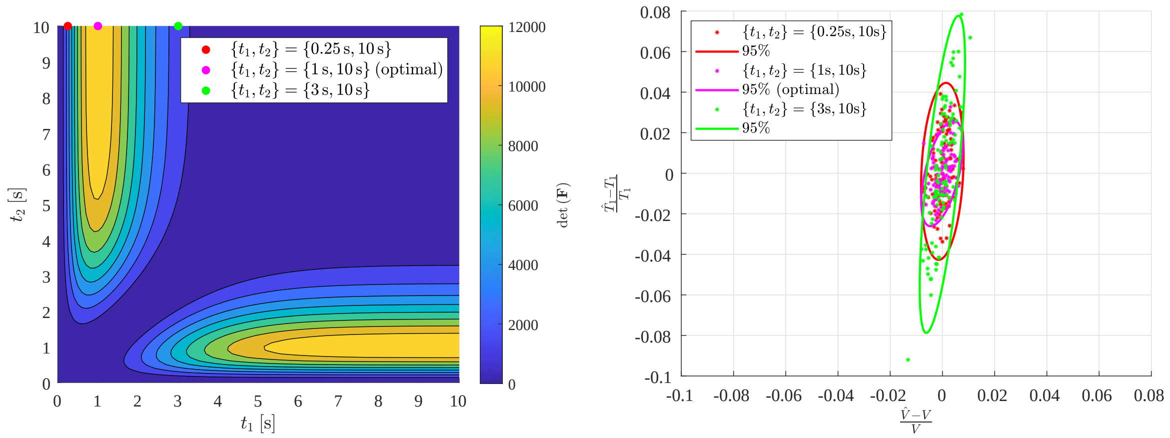
(i) $\{t_1, t_2\} = \{0.25\,\text{s}, 10\,\text{s}\}$

(ii) $\{t_1, t_2\} = \{1\,\text{s}, 10\,\text{s}\}$ (optimal)

(iii) $\{t_1, t_2\} = \{3\,\text{s}, 10\,\text{s}\}$

[1] Of course, the nominal parameters are usually not known or are subject to inaccuracies. Investigations of the effects of parameter uncertainties on the DoE task and approaches to solving this problem are discussed in Chapter 2.

As shown in Fig. 1.3(b), the parameter covariance ellipse is reduced by evaluating the model output at the optimal points in time (blue) obtained by the DoE task. A




(a) Different measurement points and Fisher information.

(b) Parameter covariance error ellipse for different measurement points, c.f. Fig. 1.3(a). For better readability, only every tenth point is shown.

Figure 1.3: 95% confidence ellipse of two-dimensional Gaussian distributed parameter covariance error.

closer look at the results shows that the uncertainties in the estimation of the gain $\hat{V}$ are in the same range of values, which makes sense because, in all three comparisons, $t_2 = 10\,\text{s}$. In contrast to this, the spread is much greater when estimating the time constant $\hat{T}_1$. Generally, the higher the information content of the respective value pair $\{t_1, t_2\}$, the lower the uncertainty of the estimated parameter, c.f., Fig. 1.3. By far, the best result is achieved by the pair of values generated by the DoE method. Thus, one can show that DoE significantly improves the parameter estimation result.

Another of this thesis's core topics, namely parameter estimation and its bias, is explained using the same introductory example. Without knowing the analytical solution of the differential equation, c.f. (1.2), the parameter estimation of CT models consists of the two main tasks:

(i) approximation or elimination of the time derivatives, resulting in a purely algebraic system of equations

(ii) determination of the parameters from the resulting algebraic system of equations, e.g., by the method of ordinary least-squares (OLS)

Without further details in this introduction, one possibility for the exact elimination of time derivatives of signals is the so-called modulating function method (MFM), which is essentially based on partial integration. One possible realization of this method, called Poisson moment functionals (PMF), can be interpreted as the convolution of a signal with impulse response of a known stable linear filter, e.g.,

$$\tilde{y}^{(i)}(t) = \frac{\mathrm{d}^i}{\mathrm{d}t}\{\tilde{y}(t)\} = \left(g^i_{F_n} * y\right)(t) \tag{1.9}$$

where $(g^i_F * y)(t)$ describes the convolution of a signal $y(t)$ with the impulse response $g^i_F(t) = \mathcal{L}^{-1}\{F^i(s)\}$ of a stable linear filter $F^i(s)$. The first order (ordinary) differential equation, equivalent to (1.1), reads

$$y(t) + T_1\dot{y}(t) = Vu(t) \tag{1.10}$$

with input $u(t) \in \mathbb{R}$, output $y(t) \in \mathbb{R}$, and the unknown (constant) parameter vector $\mathbf{p} = (T_1, V)^\top$. Using PMF, one obtains the algebraic equation for parameter estimation

$$\tilde{y}(t) = \begin{pmatrix} -\tilde{\tilde{y}}(t) & \tilde{u}(t) \end{pmatrix} \begin{pmatrix} T_1 \\ V \end{pmatrix} \tag{1.11}$$

linear in the parameters. For a single time instance $t_k$, one obtains

$$\underbrace{\tilde{y}(t_k)}_{=y_{LSQ;k}} = \underbrace{\begin{pmatrix} -\tilde{\tilde{y}}(t_k) & \tilde{u}(t_k) \end{pmatrix}}_{=\mathbf{w}_k^\top} \underbrace{\begin{pmatrix} T_1 \\ V \end{pmatrix}}_{=\mathbf{p}} . \tag{1.12}$$

Equation (1.12) is evaluated at $N$ discrete-time instances $t_k$, i.e.,

$$\underbrace{\begin{pmatrix} y_{LSQ;1} \\ y_{LSQ;2} \\ \vdots \\ y_{LSQ;k} \\ \vdots \\ y_{LSQ;N} \end{pmatrix}}_{=\mathbf{y}_{LSQ}} = \underbrace{\begin{pmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_k^\top \\ \vdots \\ \mathbf{w}_N^\top \end{pmatrix}}_{\mathbf{W}} \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n_p} \end{pmatrix}}_{\mathbf{p}} . \tag{1.13}$$

By extending the set of equations by a generalized equation error (EE)

$$\underbrace{\begin{pmatrix} y_{LSQ;1} \\ y_{LSQ;2} \\ \vdots \\ y_{LSQ;k} \\ \vdots \\ y_{LSQ;N} \end{pmatrix}}_{=\mathbf{y}_{LSQ}} = \underbrace{\begin{pmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_k^\top \\ \vdots \\ \mathbf{w}_N^\top \end{pmatrix}}_{\mathbf{W}} \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n_p} \end{pmatrix}}_{\mathbf{p}} + \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}}_{\mathbf{v}} \tag{1.14}$$

the optimal parameter estimates in the least-squares sense are obtaind by

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \,. \tag{1.15}$$

The statistical properties of the least-squares estimator are described in detail in Section 3.3.3.1. However, allow to anticipate that one of the most important properties is that the estimator yields an unbiased result when assuming a zero mean independent and identically distributed (i.i.d.) equation error $v$. Consequently, assuming a deterministic system output $y\left(t\right)$, i.e.,

$$\tilde{y}\left(t\right) = \begin{pmatrix} -\tilde{\tilde{y}}\left(t\right) & \tilde{u}\left(t\right) \end{pmatrix} \begin{pmatrix} T_1 \\ V \end{pmatrix} + v\left(t\right) \ , \quad v\left(t\right) \sim i.i.d\left(0, \sigma^2\right) \tag{1.16}$$

the estimation result is unbiased, i.e., $\hat{\mathbf{p}} = \mathbf{p}$. In case of distorted measurements, i.e.,

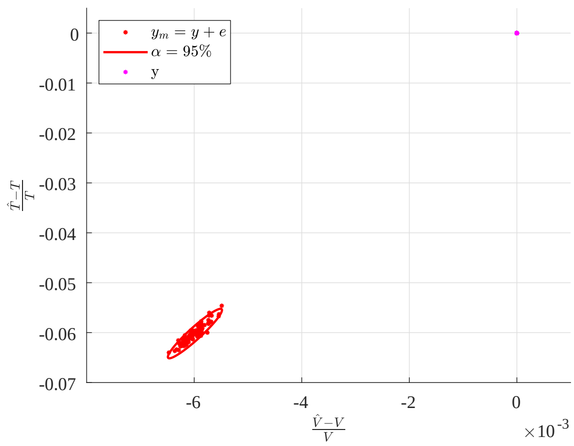
$$y_m\left(t\right) = y\left(t\right) + e\left(t\right) \ , \quad e\left(t\right) \sim i.i.d\left(0, \sigma^2\right) \tag{1.17}$$

the equation for identification (1.11) turns into

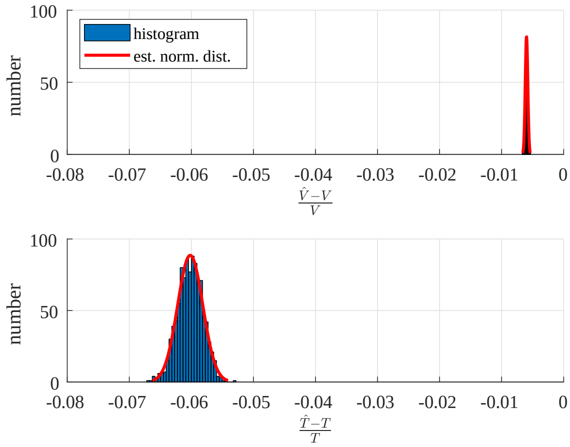
$$\tilde{y}_m\left(t\right) = \begin{pmatrix} -\tilde{\tilde{y}}_m\left(t\right) & \tilde{u}\left(t\right) \end{pmatrix} \begin{pmatrix} T \\ V \end{pmatrix} + \underbrace{\tilde{e} - T\tilde{\tilde{e}}}_{=v(t)} , \quad v\left(t\right) \nsim i.i.d\left(0, \sigma^2\right) \tag{1.18}$$

directly resulting in a biased estimation result $\hat{\mathbf{p}}$. This is because the originally independent and identically distributed (i.i.d.) measurement noise becomes a colored noise due to the PMF approach. By violating the assumptions made for the generalized equation error results in biased parameter estimation.

The simulation comparison in Fig. 1.4 clearly shows that the estimation result is biased with noisy measurements.




(a) Parameter covariance (error) ellipse, comparing ordinary least-squares (OLS) and deterministic regressor with OLS and stochastic regressor.




(b) Parameter estimation error (bias) caused by noisy output and PMF.

Figure 1.4: Comparison of the results of the OLS estimator for deterministic and stochastic regressors obtained by $y_m(t) = y(t) + e(t)$, where $e \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$ .

## 1.1 Contribution of This Thesis

The estimation of unknown model parameters is essential for many technical application areas, including control engineering. For example, an exactly parameterized model is necessary for the design of highly dynamic controllers. Online parameter estimation is also often used for specific fault monitoring applications. It is known that besides the choice of a suitable method for parameter estimation, the choice of the excitation signal influences the quality of the estimation result. For this reason, the topics of design of experiments (DoE) and parameter estimation methods for continuous-time (CT) models are discussed in this thesis.

In this work, the basic ideas of DoE for parameter estimation are reviewed, and a (graphically) illustrative relationship between the information content of signals and the quality of parameter estimation is established. For example, Fig. 1.3 shows the effect of the different measurement times to determine a first-order system's parameters.

Biased parameter estimates due to measurement noise are a common problem in system identification. Besides the shown reduction of the bias by suitable excitation signals (OID), various methods exist to reduce or eliminate the estimation bias. This work's contribution includes the detailed reappraisal of the bias estimation for continuous-time systems for Poisson moment functionals (PMF) and ordinary least-squares (OLS) from [7] and its extension to (slightly) nonlinear systems, e.g., a 1D-servo system with Coulomb static friction. Moreover, this approach to obtain asymptotically unbiased results is applied to the more general modulating function (MF) class.

Furthermore, the algorithms for optimal input design (OID) and asymptotic unbiased parameter estimation are demonstrated in measurement experiments in Chapter 4. Besides, this chapter covers another problem in the practical implementation of parameter estimation techniques: how to proceed when the system behavior leaves and re-enters the horizon covered by the identification equation. The focus is on disable and re-enable the PMF filters with appropriate initial values. Based on the slip-stick effect of a 1D-positioning system with static friction, we show these effects in a simulation experiment and offer a surprisingly simple solution for at least one particular case, validated with test-bench measurement data.

# 2 Optimal Input Design for Dynamical Systems

## 2.1 Introduction

Statistical design of experiments as a method for efficient planning and evaluation of test series was already developed in the 1920s. In 1935 Ronald A. Fisher published the first textbook with the title *The Design of Experiments* on this topic, see [4]. Nevertheless, it was not until the 1980s that the statistical design of experiments became widespread and applied worldwide.

Design of experiments is universally applicable in almost all engineering disciplines. Depending on the field of interest, this results in differences concerning the mathematical requirements. With simple experimental designs, typical questions are, for example, the necessary sample size or the distinction between real and virtual effects. Today, powerful computer simulations make it possible to create very complex experimental designs with a many factors. Of course, this also requires detailed experimental models to consider, e.g., nonlinear relationships [8]. The main goal of design of experiments is to obtain as much information as possible about the relationships between influencing variables (inputs and outputs) with as little testing effort as possible.

For dynamic systems such as those investigated here in detail, design of experiments include the choice of input and measurement ports, test signals, input-, output- and state constraints, sampling instants, etc., see, e.g., [9]. The aim is, therefore, to design experiments with high information content while adhering to application-specific constraints.

In this thesis, the focus of design of experiments is on generating optimal input signals to obtain estimates for unknown model parameters that are as accurate as possible. For this reason, it is more suitable to call the design of experiments (DoE) task optimal input design (OID). In this context, OID coincides with persistent excitation, c.f. Section 2.2.

## 2.2 Common Excitation Signals

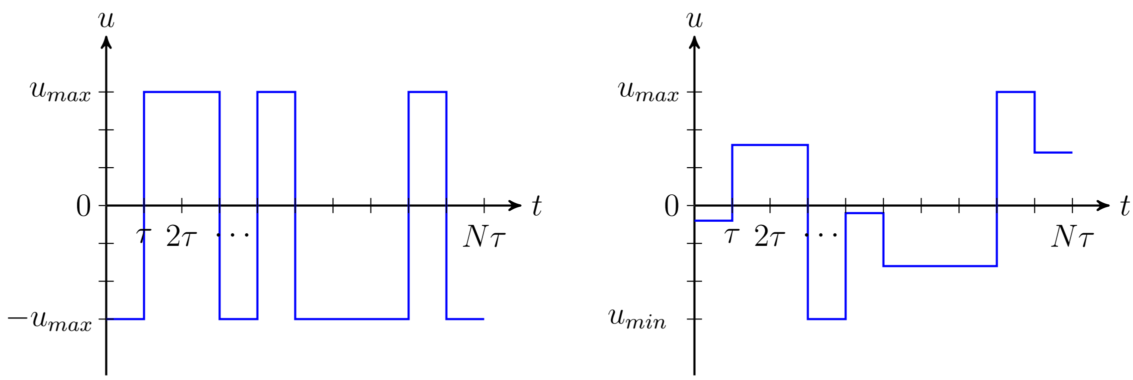
This Section briefly summarizes signal families commonly used as excitation signals to identify linear and nonlinear systems. In principle, one can differ between non-periodic signals (e.g., step, square pulse), periodic signals (e.g., sine wave, square wave), and stochastic signals (e.g., pseudorandom binary sequence (PRBS)). One can find more detailed studies on this topic, for example, in [2,3,10–12]. For instance, in [12], one can find details on the stationary and dynamic input space's different excitation coverage. According to [12], one will commonly use the following signals:

(i) pseudorandom binary sequence (PRBS) and pseudorandom multilevel sequence (PRMS)

(ii) ramps

(iii) multisine

(iv) shifted chirps




(a) PRBS. (b) pseudorandom multilevel sequence (PRMS).

Figure 2.1: PRBS and PRMS with minimum hold time $\tau$ and signal length $N$.

### 2.2.1 Pseudo Random Binary Sequence (PRBS) & Pseudo Random Multilevel Sequence (PRMS)

For the identification of linear systems, it is fundamental to excite the system over a wide frequency range. The amplitude is of secondary importance. Furthermore, binary signals provide the highest power compared with other ones with the same maximum amplitude, improving estimation accuracy for linear systems. Therefore, the so-called pseudorandom binary sequence (PRBS) is suitable as an excitation signal for identifying linear systems, which (for a sufficiently long period) has the same stochastic properties as discrete white noise. Besides minimum or maximum amplitude and signal length, the choice of minimum hold time is a design factor to affect the frequency characteristic, thus ensuring practical identifiability. For identifying, e.g., positioning systems, this must not be chosen too small to ensure that the system moves at all from a standstill.

In contrast, for nonlinear systems, a variation of the amplitudes is necessary to excite the nonlinearities, e.g., [2, 13]. An extension of PRBS allows different amplitudes to each cycle of the PRBS signal and is called pseudorandom multilevel sequence (PRMS). As stated in [14–16], PRMSs are suitable excitation signals for identifying nonlinear systems. The advantages of PRBS or PRMS signals are mainly their simple generation and applicability. As shown in [12], one disadvantage is that the user cannot precisely predefine PRBS or PRMS excitation frequencies due to its random characteristic. Of course one can use OID to distribute PRMS amplitudes better, e.g., [15].

### 2.2.2 Ramp Signals

Ramps provide similar properties to PRBS or PRMS. As pointed out in [12], against PRBS or PRMS, ramp signals have advantages concerning selectable frequencies.

### 2.2.3 Multisine Signals

A multisine signal is a sum of several harmonically related sine waves. Usually, one will use multisine signals for non-parametric system identification. The main advantage of using multisine excitation signals is that one can choose frequencies and corresponding amplitudes arbitrarily. One may find more detailed information in, e.g., [12].

### 2.2.4 Chirp Signals

Chirp signals are periodic sinusoidal signals whose frequency continuously increases decreases in a measurement period. As stated in [12], the signal power is distributed equally over the whole user-defined frequency spectrum compared to PRBS or PRMS. Advantages are the smoothness of the signal, which provides good applicability, mainly for automotive systems. The main disadvantage is the very long measurement time to cover the whole input space [12].

## 2.3 Local Design Criteria

Fundamentally, it is necessary to obtain a measure of the “goodness” of an experiment. Due to the obvious dependency between experiment and parameter estimation, it seems natural to use the expected accuracy of the parameter estimation as a measure for the quality of the experiment design. Of course, the quality of the parameter estimates depends not only on the experiment design but also on the estimator used. Hence, we assume an unbiased and efficient estimator.[1] Cramér and Rao have shown that any unbiased estimator fulfills the inequality

$$\mathrm{cov}\,(\hat{\mathbf{p}}) \geq \frac{1}{-\,\mathrm{E}[\frac{\partial^2 \ln f(\mathbf{x},\mathbf{p})}{\partial \mathbf{p}^2}\Big|_{\mathbf{p}_0}]} \tag{2.1}$$

where

$$\mathrm{cov}\,(\hat{\mathbf{p}}) = \mathrm{E}[(\hat{\mathbf{p}} - \mathbf{p})\,(\hat{\mathbf{p}} - \mathbf{p})^\top] \tag{2.2}$$

is the parameter covariance matrix. The nominal parameter vector is $\mathbf{p}_0$, the estimated vector is denoted by $\hat{\mathbf{p}}$, and $f$ is the probability density function (PDF), see, e.g. [9,17]. For finding a proper design criterion, it is reasonable to assume an unbiased and efficient estimator, whereby inequality (2.1) results in the Cramér-Rao lower bound (CRLB)

$$\mathrm{CRLB} = \frac{1}{-\,\mathrm{E}[\frac{\partial^2 \ln f(\mathbf{x},\mathbf{p})}{\partial \mathbf{p} \partial \mathbf{p}^\top}]} \tag{2.3}$$

where the denominator

$$\mathbf{F}\,(\mathbf{p}) = -\,\mathrm{E}[\frac{\partial^2 \ln f\,(\mathbf{x},\mathbf{p})}{\partial \mathbf{p} \partial \mathbf{p}^\top}] \tag{2.4}$$

is denoted as Fisher information.[2] Consequently, maximizing the Fisher information results in a decreasing CRLB (2.3), and therefore more accurate parameter estimates are possible, provided that a suitable estimator is used.

[1]The statistical properties of estimators are summarized in Section 3.2.

[2]Cramér-Rao lower bound (CRLB) provides information about the best estimation values, i.e., lowest parameter variance, one can expect.

## 2.4 Fisher Information Matrix

We consider multiple-input multiple-output (MIMO) continuous-time (CT) systems

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}\left(\mathbf{x}, \mathbf{u}, \mathbf{p}\right) \\ \mathbf{y} &= \mathbf{h}\left(\mathbf{x}, \mathbf{u}, \mathbf{p}\right) \end{aligned} \tag{2.5}$$

with state vector $\mathbf{x} \in \mathbb{R}^n$, output vector $\mathbf{y} \in \mathbb{R}^{n_y}$, constant parameter vector $\mathbf{p} \in \mathbb{R}^{n_p}$, input vector $\mathbf{u} \in \mathbb{R}^{n_u}$, and vector fields $\mathbf{f}(\cdot), \mathbf{h}(\cdot)$. Usually, these continuous-time (CT) systems are evaluated at specific discrete points in time $t_k$, $1 \le k \le N$, where $N$ denotes the total number of measured samples. Deviations of the measurement data from the model, caused by model- or measurement errors, are considered additive output noise. The continuous-time model is sampled with a particular sample time $T$, resulting in discrete-time instances $t_k = kT$. Furthermore, we assume noise corrupted samples

$$y_{m;i}\left(t_k\right) = y_i\left(t_k\right) + e_i\left(t_k\right) \ , \quad 1 \le i \le n_y \ , \quad 1 \le k \le N \ . \tag{2.6}$$

For the sake of brevity, we introduce $y_{m;i}\left(t_k\right) = y_{m;i,k}$ and write

$$y_{m;i,k} = y_{i,k} + e_{i,k} \ , \quad 1 \le i \le n_y \ , \quad 1 \le k \le N \tag{2.7}$$

where $\mathbf{y}_{m;i} = \left(y_{m;i,1}, y_{m;i,2}, \ldots, y_{m;i,N}\right)^\top \in \mathbb{R}^N$ is a vector containing $N$ samples of the $i$-th measured system output. Note that by inserting the deterministic model in the stochastic description of the data, the measurement equation must be written precisely as

$$y_{m;i,k} = y_{i,k}\left(\mathbf{p}_0\right) + e_{i,k} \tag{2.8}$$

where the vector $\mathbf{p}_0$ describes the nominal parameter values. For better readability, in this thesis, one uses $\mathbf{p}$ instead of $\mathbf{p}_0$ as long as there is no risk of confusion. The $i$-th deterministic model output is $y_i\left(t_k\right) = y_{i,k}$ with the corresponding noise sequence $e_{i,k}$.

The error is assumed to be Gaussian white noise, consisting of $n_y$ sequences of independent random variables, i.e., $e_{i,k}$ , $1 \leq i \leq n_y$, each normally distributed, with zero mean and finite variance (see Section A.3.7).

$$
\begin{aligned}
\mathrm{E}[e_{i,k}] &= 0, \quad 1 \leq i \leq n_y \\
\mathrm{E}[e_{i,k} e_{i+j,k}] &= \begin{cases} \sigma_i^2, & j = 0 \\ 0, & j \neq 0 \end{cases} \\
\mathrm{E}[e_{i,k} e_{i,k+l}] &= \begin{cases} \sigma_i^2, & l = 0 \\ 0, & l \neq 0 \end{cases} \\
e_{i,k} &\sim \mathcal{N}(0, \sigma_i^2), \quad 1 \leq i \leq n_y \,.
\end{aligned}
\tag{2.9}
$$

With

$$
\mathbf{e}_i = \left(e_{i,1}, e_{i,2}, \ldots, e_{i,N}\right) \in \mathbb{R}^N \tag{2.10}
$$

and

$$
\mathbf{e} = \left(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{n_y}\right)^\top \in \mathbb{R}^{n_y N} \tag{2.11}
$$

the probability density function (PDF) for the multivariate normally distributed error $\mathbf{e} \sim \mathcal{N}_{n_y N}\left(\mathbf{0}, \mathbf{\Sigma}\right)$ reads

$$
f\left(\mathbf{e}\right) = \frac{1}{\sqrt{(2\pi)^{n_y N} \det\left(\mathbf{\Sigma}\right)}} e^{-\frac{1}{2}\mathbf{e}^\top (\mathbf{\Sigma})^{-1}\mathbf{e}} \tag{2.12}
$$

where $\mathbf{\Sigma} \in \mathbb{R}^{n_y N \times n_y N}$ is the error covariance matrix. Considering stochastically independent errors $\mathbf{e}_i$, one obtains

$$
f\left(\mathbf{e}\right) = (2\pi)^{-\frac{n_y N}{2}} \left(\prod_{i=1}^{n_y} \sigma_i^2\right)^{-\frac{N}{2}} e^{-\frac{1}{2}\sum_{k=1}^{N}\sum_{i=1}^{n_y}\frac{e_{i,k}^2}{\sigma_i^2}} \,. \tag{2.13}
$$

Assuming that the estimated model parameters $\hat{\mathbf{p}}$ are very close to the nominal parameter values, (2.13) can be approximated by

$$
f\left(\mathbf{e}\right) \approx f\left(\mathbf{p}_0\right) = (2\pi)^{-\frac{n_y N}{2}} \left(\prod_{i=1}^{n_y} \sigma_i^2\right)^{-\frac{N}{2}} e^{-\frac{1}{2}\sum_{k=1}^{N}\sum_{i=1}^{n_y}\frac{\left(y_{i,k}(\mathbf{p}_0) - y_{m;i,k}\right)^2}{\sigma_i^2}} \,. \tag{2.14}
$$

Inserting

$$\frac{\partial^2 \ln\left(f\left(\mathbf{p}_0\right)\right)}{\partial \mathbf{p} \partial \mathbf{p}^\top} = -\sum_{k=1}^{N}\sum_{i=1}^{n_y}\frac{1}{\sigma_i^2}\left(\left.\frac{\partial y_i}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k}^\top \left.\frac{\partial y_i}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k} + e_{i,k}\left.\frac{\partial y_i}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k}\right) \tag{2.15}$$

in (2.4) and taking into account that one uses expectations, using the noise assumption $\mathrm{E}[e_{i,k}] = 0$ (2.9) and $\mathrm{E}[\hat{\mathbf{p}}] = \mathbf{p}_0$, one obtains

$$\mathbf{F}\left(\mathbf{u}, \mathbf{p}_0\right) = -\frac{\partial^2 \ln\left(f\left(\mathbf{p}_0\right)\right)}{\partial \mathbf{p} \partial \mathbf{p}^\top} = \sum_{k=1}^{N}\sum_{i=1}^{n_y}\frac{1}{\sigma_i^2}\left(\left.\frac{\partial y_i}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k}^\top \left.\frac{\partial y_i}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k}\right) . \tag{2.16}$$

In a more compact vector form

$$\mathbf{F}(\mathbf{u}, \mathbf{p}_0) = \sum_{k=1}^{N}\left.\frac{\partial \mathbf{y}}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k}^\top \left(\boldsymbol{\sigma}^2\right)^{-1}\left.\frac{\partial \mathbf{y}}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k} , \quad \mathbf{F}(\mathbf{u}, \mathbf{p}_0) \in \mathbb{R}^{n_p \times n_p} \tag{2.17}$$

where

$$\left.\frac{\partial \mathbf{y}}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k} = \begin{pmatrix} \left.\frac{\partial y_1}{\partial p_1}\right|_{\mathbf{p}_0,t_k} & \cdots & \left.\frac{\partial y_1}{\partial p_{n_p}}\right|_{\mathbf{p}_0,t_k} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial y_{n_y}}{\partial p_1}\right|_{\mathbf{p}_0,t_k} & \cdots & \left.\frac{\partial y_{n_y}}{\partial p_{n_p}}\right|_{\mathbf{p}_0,t_k} \end{pmatrix} , \quad \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \in \mathbb{R}^{n_y \times n_p} \tag{2.18}$$

and

$$\boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ 0 & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n_y}^2 \end{pmatrix} , \quad \boldsymbol{\sigma}^2 \in \mathbb{R}^{n_y \times n_y} . \tag{2.19}$$

assuming stochastically independent outputs $y_i$, c.f. [9, 17]. If the assumption that the errors $e_i$ are stochastically independent is not valid, the covariance matrix diagonal entries are unequal to zero, e.g. [18], resulting in

$$\boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n_y}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2n_y}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_y1}^2 & \sigma_{n_y2}^2 & \cdots & \sigma_{n_y}^2 \end{pmatrix} , \quad \boldsymbol{\sigma}^2 \in \mathbb{R}^{n_y \times n_y} . \tag{2.20}$$

For a large number of samples $N$, it is more convenient to introduce the so-called average Fisher matrix as suggested in, e.g., [9, 19, 20].

$$\bar{\mathbf{F}}\left(\mathbf{u},\mathbf{p}_0\right)=\frac{1}{N}\sum_{k=1}^{N}\left.\frac{\partial \mathbf{y}}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k}^{\top}\left(\boldsymbol{\sigma}^2\right)^{-1}\left.\frac{\partial \mathbf{y}}{\partial \mathbf{p}}\right|_{\mathbf{p}_0,t_k} \tag{2.21}$$

As already mentioned in the previous section, the aim is to generate experiments with high information content. This intention maximizes the FIM (2.17) or, equally, minimizes the inverse FIM (Cramér-Rao matrix). The Cramér-Rao matrix provides a lower bound of the parameter covariance matrix $\mathbf{P}$, i.e.,

$$\mathbf{P}\geq\mathbf{F}^{-1}\left(\mathbf{u},\mathbf{p}_0\right)\ . \tag{2.22}$$

### 2.4.1 Standard Metrics for the “Size” of the Fisher Information Matrix

To push down the lower bound of the parameter covariance matrix, the FIM must be maximized. Therefore, for optimization purposes, it is necessary to derive scalar functions from the Fisher matrix. In the literature, various real-valued functions have been suggested as “suitable” metrics for the size of Fisher matrix or the parameter covariance matrix, e.g., [18, 21]. The most commonly used criteria are the D-criterion, A-criterion, and E-criterion.[3,4]

- D-criterion: maximize the determinant of the Fisher matrix, which is equivalent to minimization of the volume of the uncertainty ellipsoid.

  $$D_{\mathrm{OPT}}=\max_{\mathbf{u}}\det\left(\mathbf{F}\left(\mathbf{u},\mathbf{p}\right)\right)=\min_{\mathbf{u}}\det\left(\mathbf{P}\left(\mathbf{u},\mathbf{p}\right)\right) \tag{2.23}$$

  Advantages of the most used criterion [20] are the easy geometric interpretation [22] and the (theoretical) invariance to parameter scaling and linear transformations of the output [18, 23]. The drawback is that the D-criterion tends to give disproportionate weight to model parameters to which the model is sensitive. This results in a decreasing variance of this parameter, but the uncertainty of other parameters remains quite large, see [24].

[3] For simplicity, the different norms are written down here for the original Fisher matrix according to (2.17). Of course, they are also valid for the average Fisher matrix (2.21) and the scaled Fisher matrix (2.27).

[4] For better readability of this thesis, one uses $\mathbf{p}$ instead of $\mathbf{p}_0$ from this point on, unless there is a risk of confusion.

- A-criterion: maximize the trace of the Fisher matrix, which is equivalent to minimization of the average variance of the parameters.

$$A_{\mathrm{OPT}} = \max_{\mathbf{u}} \mathrm{Tr}\left(\mathbf{F}\left(\mathbf{u},\mathbf{p}\right)\right) = \min_{\mathbf{u}} \mathrm{Tr}\left(\mathbf{P}\left(\mathbf{u},\mathbf{p}\right)\right) \tag{2.24}$$

A disadvantage of the A-criterion is that it may lead to non-informative experiments in a high correlation of the model parameters [24–26]. This is caused by omitting the minor diagonal elements of the matrix.

- E-criterion: maximize the smallest eigenvalue of the Fisher-matrix or minimizes the largest eigenvalue of the parameter covariance matrix.

$$E_{\mathrm{OPT}} = \max_{\mathbf{u}} \lambda_{\min}\left(\mathbf{F}\left(\mathbf{u},\mathbf{p}\right)\right) = \min_{\mathbf{u}} \lambda_{\max}\left(\mathbf{P}\left(\mathbf{u},\mathbf{p}\right)\right) \tag{2.25}$$

- $E_{\mathrm{mod}}$-criterion: minimize the ratio between the largest and smallest eigenvalue of the Fisher matrix.

$$E_{\mathrm{OPT}}^{mod} = \min_{\mathbf{u}} \frac{\lambda_{\max}\left(\mathbf{F}\left(\mathbf{u},\mathbf{p}\right)\right)}{\lambda_{\min}\left(\mathbf{F}\left(\mathbf{u},\mathbf{p}\right)\right)} \tag{2.26}$$

A geometric interpretation of the most important criteria for the two-parameter case is shown in Fig. 2.2. The D-criterion corresponds with the volume of the uncertainty ellipsoid (confidence region). The A-criterion is a measure of the surrounding box of the uncertainty ellipsoid, and the E-criterion is equivalent to the length of the major axis of the ellipsoid.

Other, less common criteria include:

- G-criterion
- L-criterion
- C-criterion
- $D_s$-criterion

One can find more detailed information about the different criteria, e.g., [20, 23].

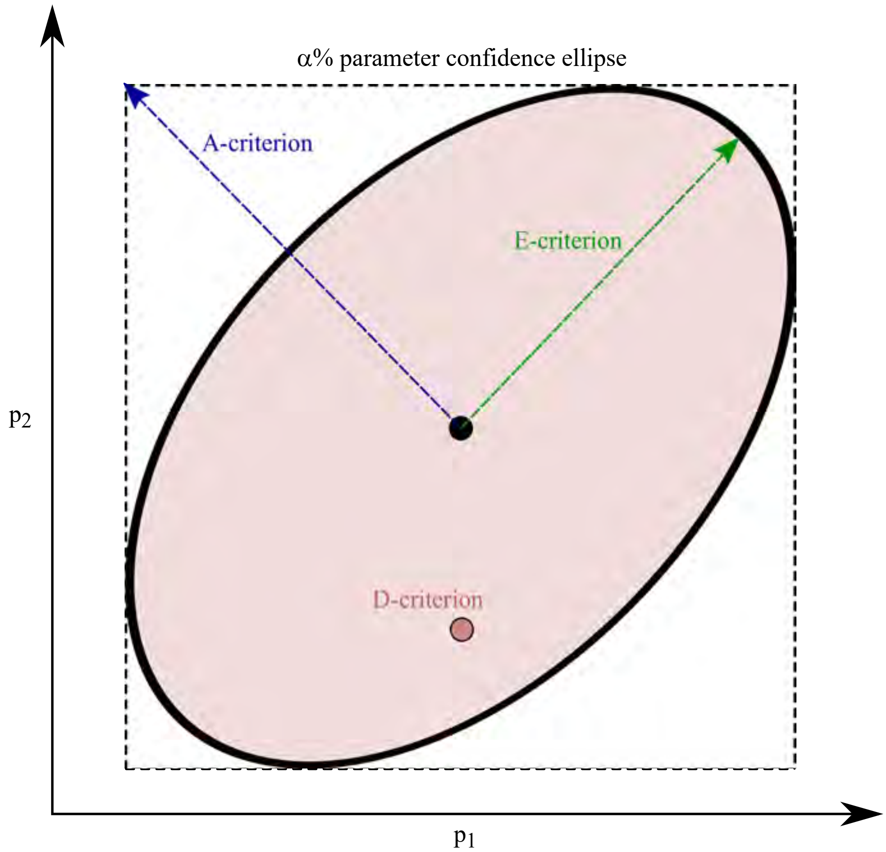



Figure 2.2: Geometric interpretation of the most important criteria for optimal input design in case of two parameters. The ellipsoid represents the $\alpha\%$ confidence region of the estimated parameters, where usually $\alpha = 90 - 95\%$.

## 2.4.2 Miscellaneous Topics on the Fisher Matrix

### 2.4.2.1 Initial Guess of Model Parameters

As can be seen in (2.17) or (2.18), the Fisher matrix depends, among other things, on the unknown nominal model parameters to be estimated later on. To run an OID iteration, one uses the best available parameter values (initial guess) for building the Fisher matrix. Such a strategy usually claims an iterative design procedure if the initial parameter guesses are unreliable.

### 2.4.2.2 Parameter Analysis and Parameter Estimate Error

One can use the Fisher matrix to perform a parameter analysis. In concrete terms, this means that one can identify model parameters that can be estimated with given maximum variance. Conversely, this means that parameters with a more considerable variance can not be estimated with the desired accuracy using the available measurement data. It is also possible to detect (local) problems, e.g., loss of structural identifiability. As already mentioned, the diagonal elements of the inverse of the Fisher matrix represent a lower limit for the following parameter estimation variances. Consequently, a singular Fisher matrix (at least one eigenvalue is zero) means that there exist parameters or parameter combinations that cannot be estimated with the measurement data available, see [27]. For the parameter analysis, one can use the parameter values known at that time. The following parameter estimation provides preliminary optimal parameters. This process of parameter analysis and estimation is iterated until convergence is achieved. In [17], it is suggested to use the Fisher matrix with scaled sensitivities to compute the lower bound of the parameter estimation error covariance matrix, i.e.,

$$\tilde{\mathbf{F}} = \sum_{k=1}^{N} \left( \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \mathbf{\Lambda}_p \right)\Bigg|_{t_k}^{\top} \left(\boldsymbol{\sigma}^2\right)^{-1} \left( \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \mathbf{\Lambda}_p \right)\Bigg|_{t_k} \tag{2.27}$$

where

$$\mathbf{\Lambda}_p = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ 0 & & \ddots & 0 \\ 0 & \cdots & 0 & p_{n_p} \end{pmatrix}, \quad \mathbf{\Lambda}_p \in \mathbb{R}^{n_p \times n_p} . \tag{2.28}$$

Assuming constant parameters, one can rewrite (2.27) to

$$\begin{aligned} \tilde{\mathbf{F}} &= \mathbf{\Lambda}_p \sum_{k=1}^{N} \left( \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \right)\Bigg|_{t_k}^{\top} \left(\boldsymbol{\sigma}^2\right)^{-1} \left( \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \right)\Bigg|_{t_k} \mathbf{\Lambda}_p \\ &= \mathbf{\Lambda}_p \mathbf{F} \mathbf{\Lambda}_p . \end{aligned} \tag{2.29}$$

For details, see [17].

#### 2.4.2.3 Informative Experiments / (Local) Identifiability

Provided that the system is structurally identifiable (Section 3.1.3), in [28], a sufficient condition for local identifiability is given by

$$\det \left( \frac{\partial \mathbf{y}}{\partial \mathbf{p}}\Bigg|^{\top} \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \right) \neq 0 . \tag{2.30}$$

Consequently, a non-degenerated or informative experiment satisfies

$$\det \left( \mathbf{F} \left( \mathbf{u}, \mathbf{p} \right) \right) \neq 0 \tag{2.31}$$

and ensures local identifiability of the model parameters [19]. For symmetrical matrices, it applies

$$\det\left(\mathbf{F}\left(\mathbf{u},\mathbf{p}\right)\right)=\prod_{i=1}^{n_p}\lambda_i\,. \tag{2.32}$$

Consequently, the statement about the identifiability in this section is identical to the one made in Section 2.4.2.2 based on the Fisher matrix's eigenvalues.

### 2.4.3 Sensitivity Ordinary Differential Equations

Unfortunately, in general, it is not possible to derive an analytical solution for the sensitivities. This only works for quite simple models. Therefore, the sensitivities for more general models have to be determined numerically. The local sensitivity of the solution to parameters can be calculated simultaneously with the model (2.5) using the sensitivity ordinary differential equation (ODE)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{S}&=\frac{\partial\mathbf{f}\left(\mathbf{x},\mathbf{p},\mathbf{u},t\right)}{\partial\mathbf{x}}\mathbf{S}+\frac{\partial\mathbf{f}\left(\mathbf{x},\mathbf{p},\mathbf{u},t\right)}{\partial\mathbf{p}}=\mathbf{JS}+\mathbf{M}\\ \mathbf{S}\left(0\right)&=\frac{\partial\mathbf{x}_0\left(\mathbf{p}\right)}{\partial\mathbf{p}}\end{aligned} \tag{2.33}$$

where

$$\mathbf{S}=\begin{pmatrix}\frac{\partial x_1(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_1} & \cdots & \frac{\partial x_1(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_{n_p}}\\ \vdots & \ddots & \vdots\\ \frac{\partial x_n(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_1} & \cdots & \frac{\partial x_n(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_{n_p}}\end{pmatrix},\ \mathbf{S}\in\mathbb{R}^{n\times n_p} \tag{2.34}$$

is the sensitivity matrix,

$$\mathbf{J}=\begin{pmatrix}\frac{\partial f_1(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial f_n(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial x_1} & \cdots & \frac{\partial f_n(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial x_n}\end{pmatrix},\ \mathbf{J}\in\mathbb{R}^{n\times n} \tag{2.35}$$

is the Jacobian, and

$$\mathbf{M}=\begin{pmatrix}\frac{\partial f_1(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_1} & \cdots & \frac{\partial f_1(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_{n_p}}\\ \vdots & \ddots & \vdots\\ \frac{\partial f_n(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_1} & \cdots & \frac{\partial f_n(\mathbf{x},\mathbf{p},\mathbf{u},t)}{\partial p_{n_p}}\end{pmatrix},\ \mathbf{M}\in\mathbb{R}^{n\times n_p} \tag{2.36}$$

is the parameter derivatives matrix. For multiple-input multiple-output (MIMO) linear time-invariant (LTI) systems

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}\end{aligned} \tag{2.37}$$

the sensitivity equation (2.33) simplifies to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{S} &= \mathbf{A}\mathbf{S} + \mathbf{M} \\ \mathbf{S}(0) &= \frac{\partial \mathbf{x}_0(\mathbf{p})}{\partial \mathbf{p}}\end{aligned} \tag{2.38}$$

where $\mathbf{x}_0 = \mathbf{x}(0)$.

Depending on which state variables are measured outputs, the output sensitivities $\mathbf{S}_y \in \mathbb{R}^{n_y \times n_p}$ are individual rows or combinations from $\mathbf{S}$, used to build the Fisher matrix, i.e.,

$$\mathbf{F} = \sum_{k=1}^{N} \mathbf{S}_y\Big|_{t_k}^{\top} \left(\boldsymbol{\sigma}^2\right)^{-1} \mathbf{S}_y\Big|_{t_k}, \quad \mathbf{F} \in \mathbb{R}^{n_p \times n_p}. \tag{2.39}$$

### 2.4.4 Global Parameter Sensitivities

The use of local sensitivities is quasi-standard for analyzing the effects of parameter variations [29]. However, it is disadvantageous that these local sensitivities are strictly speaking only valid in the immediate neighborhood of a reference point, which is usually not known in practical applications. Moreover, the idea of local sensitivities is based on linearization principles. Hence, nonlinear models are linearized around, possibly unknown, reference parameter values, and thus nonlinear effects of models are lost. One possibility to circumvent this problem is the use of so-called global sensitivities, shown in, e.g., [29]. This concept is based on the idea that parameters and model outputs are understood as random variables, and the respective contribution to the total variance of the output is quantified [29–32]. One may use the first-order Sobol sensitivity index

$$S_i^G(t) = \frac{\sigma_i^2\left(\mathop{\mathrm{E}}_{-i}\left[y(t)\,|p_i\right]\right)}{\sigma^2(y(t))} \tag{2.40}$$

to analyze parameter sensitivities and parameter interactions, where $\sigma_i^2\left(\mathop{\mathrm{E}}_{-i}\left[y(t)\,|p_i\right]\right)$ represents the contribution of the $i-$th parameter to the total variance $\sigma^2(y(t))$. The local design measures introduced in Section 2.4.1 are generalized to hold global sensitivities, e.g., [29]. DoE or OID approaches using global sensitivities instead of local sensitivities are often used in chemistry. The use of global sensitivities is intentionally omitted in this work. More in-depth literature on this topic can be found, for example, in [29, 33, 34].

## 2.5 Investigations on the Subject of Optimal Input Design Using a First Order Transfer Function

This section shows the potential of DoE methods for dynamic systems based on a straightforward, analytically solvable system. We consider a simple single-input single-output (SISO) linear time-invariant system

$$G\left(s\right) = \frac{\hat{y}\left(s\right)}{\hat{u}\left(s\right)} = \frac{V}{1 + sT_1} \tag{2.41}$$

with its parameters $V$ (gain) and $T_1$ (time constant), which should be determined best possible. For the following simulation experiments, nominal model parameters $V = 3$ and $T_1 = 1$ are set. The system is excited with a unit step, for which the system response can be analytically computed. To determine the unknown parameters, at least two independent equations are necessary, evaluating the output at two different time instances. Therefore, the DoE task consists of finding two optimal time instances where the system response is evaluated.

The analytical solution $y\left(t\right)$ for the step response $u\left(t\right) = \sigma\left(t\right)$ can be written as

$$y\left(t\right) = \mathcal{L}^{-1}\left(\frac{1}{s}G\left(s\right)\right) = V\left(1 - e^{-\frac{t}{T_1}}\right) . \tag{2.42}$$

The sensitivities are calculated by partial derivation of the output according to the parameters (2.18) and reveal

$$\begin{aligned} s_1\left(t\right) &= \left.\frac{\partial y}{\partial V}\right|_{\mathbf{p}_0} = 1 - e^{-\frac{t}{T_1}} \\ s_2\left(t\right) &= \left.\frac{\partial y}{\partial T_1}\right|_{\mathbf{p}_0} = -\frac{tV}{T_1^2}e^{-\frac{t}{T_1}} . \end{aligned} \tag{2.43}$$

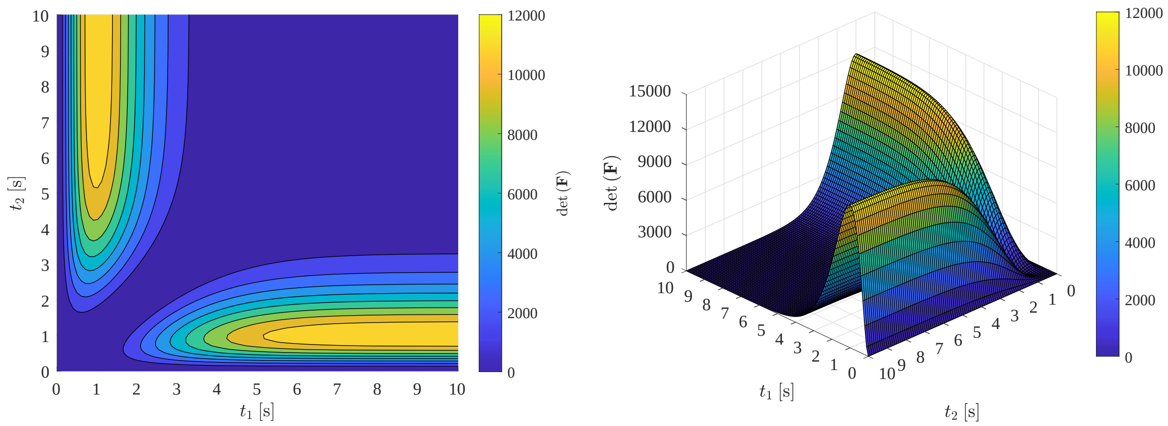
Inserting in (2.17) and evaluating at $N = 2$ time instances $t_k$ results in

$$\begin{aligned}\mathbf{F}(u,\mathbf{p}_0)|_{u=\sigma(t)} &= \sum_{k=1}^{2}\frac{1}{\sigma^2}\begin{pmatrix}\left(\frac{\partial y}{\partial V}\right)^2\Big|_{\mathbf{p}_0,t_k} & \left(\frac{\partial y}{\partial V}\frac{\partial y}{\partial T_1}\right)\Big|_{\mathbf{p}_0,t_k}\\ \left(\frac{\partial y}{\partial T_1}\frac{\partial y}{\partial V}\right)\Big|_{\mathbf{p}_0,t_k} & \left(\frac{\partial y}{\partial T_1}\right)^2\Big|_{\mathbf{p}_0,t_k}\end{pmatrix}\\ &= \frac{1}{\sigma^2}\sum_{k=1}^{2}\begin{pmatrix}\left(1-e^{-t_k/T_1}\right)^2 & \left(1-e^{-t_k/T_1}\right)\left(-\frac{t_kV}{T_1^2}e^{-\frac{t_k}{T_1}}\right)\\ \left(1-e^{-t_k/T_1}\right)\left(-\frac{t_kV}{T_1^2}e^{-\frac{t_k}{T_1}}\right) & \left(-\frac{t_kV}{T_1^2}e^{-\frac{t_k}{T_1}}\right)^2\end{pmatrix}.\end{aligned} \tag{2.44}$$

For optimization purposes, the D-criterion, i.e., the determinant of the Fisher matrix, is used as the objective function, and one ends up with the constrained optimization problem

$$\begin{aligned}&\max_{t_1,t_2}\det\left(\mathbf{F}\left(u,\mathbf{p}_0\right)\right)\\ &t_1\in[0,10]\\ &t_2\in[0,10]\ .\end{aligned} \tag{2.45}$$

Figure 2.3 shows the information content, $\det\left(\mathbf{F}\left(u,\mathbf{p}_0\right)\right)$, as a function of the time instances $t_1$ and $t_2$. It is easy to see that the maximum information, and thus the optimum measurement times, are $\{t_1,t_2\} = \{1\,\text{s},10\,\text{s}\}$ or $\{t_1,t_2\} = \{10\,\text{s},1\,\text{s}\}$, respectively. This result seems quite plausible since it means that the gain factor $V$ can be




(a) Contour plot. (b) Three-dimensional surface plot.

Figure 2.3: Optimal measurement points $t_1$ and $t_2$ for estimating the model parameters $V$ and $T_1$ using D-criterion for evaluating the Fisher matrix with $\sigma = 0.1$.

determined best in the steady-state range and the time constant $T_1$ within the range of large slope.

### 2.5.1 Parameter Estimation & Monte Carlo Simulation

We assume a noise-affected system output for parameter estimation, where the noise is supposed to be normally distributed with zero mean and variance $\sigma^2 = 0.01$. Figure. 2.4 shows one realization of the model input and output. One can determine

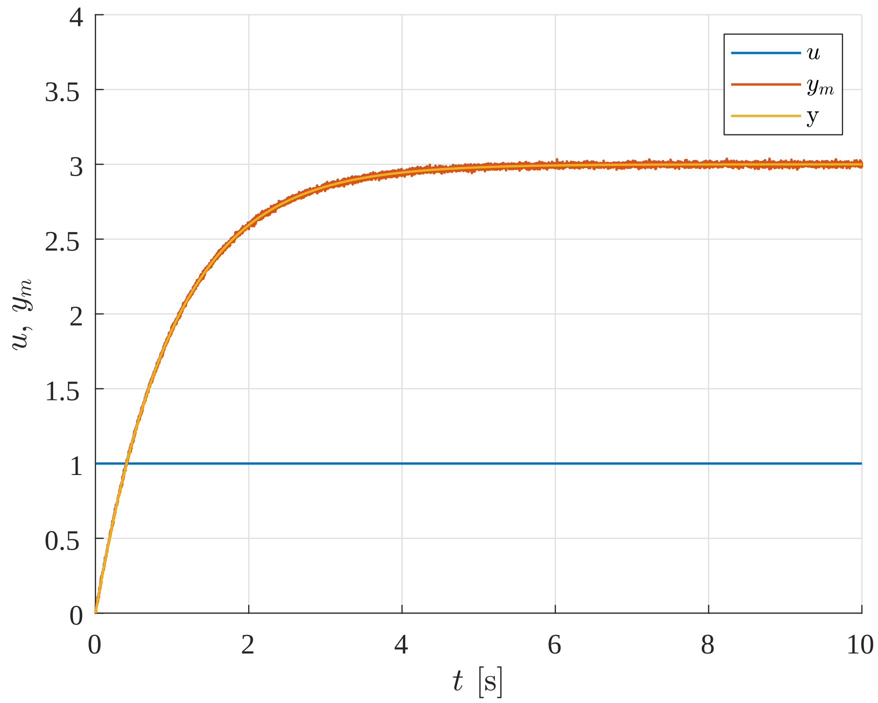



Figure 2.4: Step response with additive output noise. Noisy output $y_m(t) = y(t) + e(t)$, where $e \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$.

the first-order system's parameters with two measurements, $y(t_1)$, $y(t_2)$, using the analytical solution (2.42) with

$$\begin{aligned} \hat{V} &= \lim_{t_2 \to \infty} y(t_2) \\ \hat{T}_1 &= -\frac{t_1}{\ln\left(1 - \frac{y(t_1)}{\hat{V}}\right)} \,. \end{aligned} \tag{2.46}$$

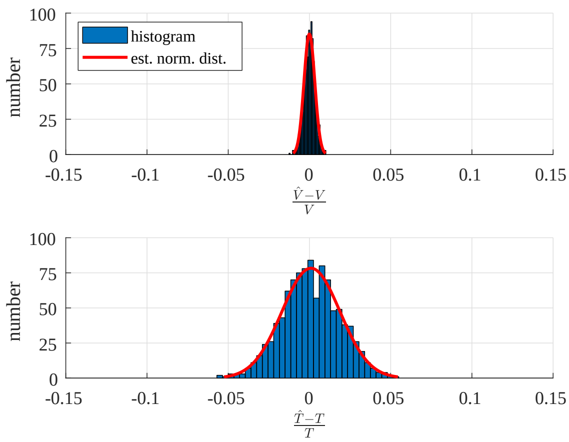
To determine whether the estimated parameter values are affected by choice of evaluation times, we perform a thousand simulation runs and evaluate them at three different pairs of time values. More precisely:

(i) $\{t_1, t_2\} = \{0.25\,\text{s}, 10\,\text{s}\}$

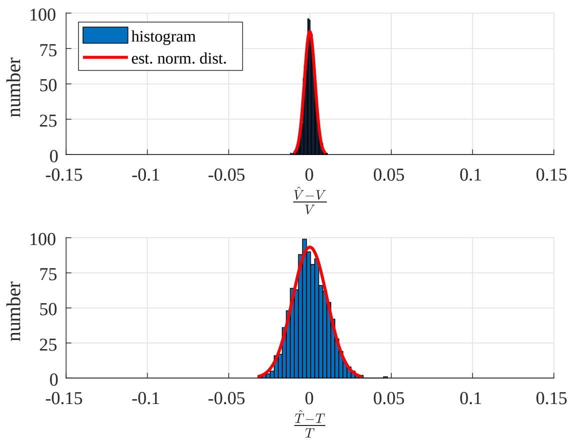
(ii) $\{t_1, t_2\} = \{1\,\text{s}, 10\,\text{s}\}$ (optimal)

(iii) $\{t_1, t_2\} = \{3\,\text{s}, 10\,\text{s}\}$

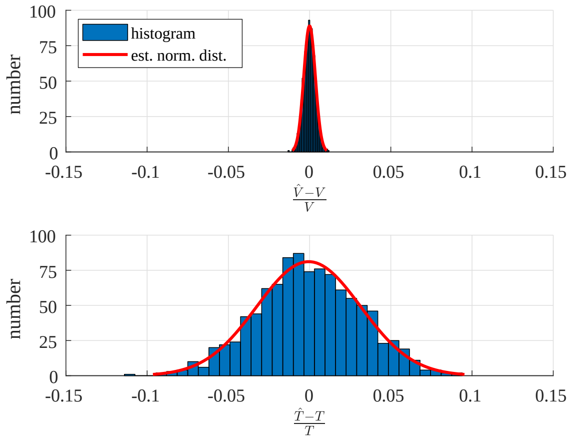
Figure 2.5 shows the respective estimation errors and corresponding fitted normal distribution histograms.




(a) $t_1 = 0.25\,\mathrm{s}$ and $t_2 = 10\,\mathrm{s}$



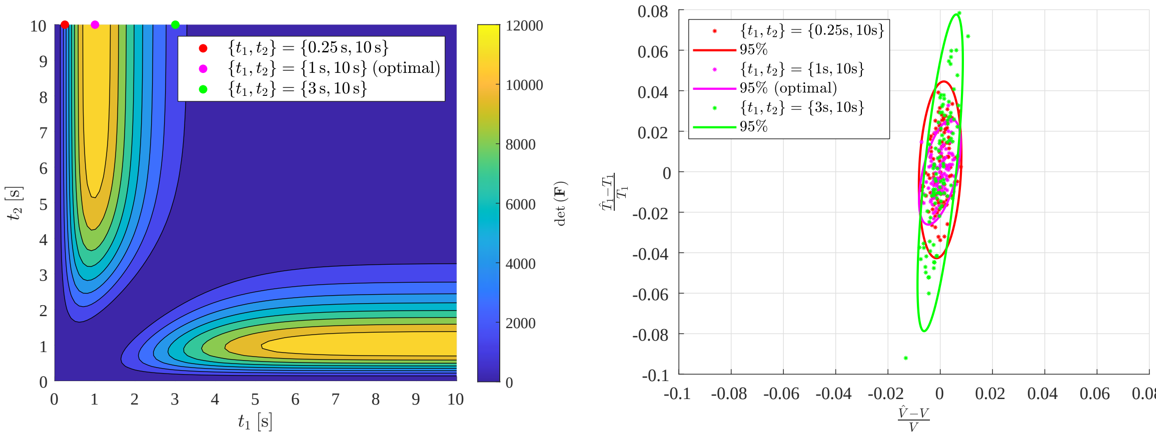

(b) $t_1 = 1\,\mathrm{s}$ and $t_2 = 10\,\mathrm{s}$ (optimal)




(c) $t_1 = 3\,\mathrm{s}$ and $t_2 = 10\,\mathrm{s}$

Figure 2.5: Parameter estimation error in dependence of evaluation times $\{t_1, t_2\}$.

As shown in Fig. 2.6(b), the parameter covariance ellipse is reduced by evaluating the model output at the optimal points in time, calculated by the Fisher matrix. A smaller covariance matrix means that the parameters estimate results are closer to the nominal values. Thus, one can show that DoE significantly improves the ensuing parameter estimation.




(a) Different measurement points and Fisher information.

(b) Parameter covariance error ellipse for different measurement points, c.f. Fig. 2.6(a). For reasons of better clarity, only every tenth data point is visualized.

Figure 2.6: 95% confidence ellipse of two-dimensional gausssian distributed parameter covariance error.

| | (a) | | (b) optimal | | (c) | |
|---|---|---|---|---|---|---|
| | $\Delta V$ | $\Delta T_1$ | $\Delta V$ | $\Delta T_1$ | $\Delta V$ | $\Delta T_1$ |
| $\hat{\mu}$ | $-9.55 \times 10^{-5}$ | $-1.203 \times 10^{-3}$ | $-9.55 \times 10^{-5}$ | $-1.11 \times 10^{-4}$ | $-9.55 \times 10^{-5}$ | $-1.4 \times 10^{-3}$ |
| $\hat{\sigma}$ | $1.01 \times 10^{-2}$ | $1.722 \times 10^{-2}$ | $1.01 \times 10^{-2}$ | $1.064 \times 10^{-2}$ | $1.01 \times 10^{-2}$ | $3.057 \times 10^{-2}$ |

Table 2.1: Estimation errors $\Delta\hat{V} = \hat{V} - V$ and $\Delta\hat{T}_1 = \hat{T}_1 - T_1$ evaluated at different measurement points.

## 2.6 Investigations on the Subject of Optimal Input Design Using a 1D-Servo System

### 2.6.1 Linear Model (2 Parameters)

Considering the linear 1D-servo system

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\frac{d}{m} \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{m} \end{pmatrix} u \tag{2.47}$$

with state vector $\mathbf{x} = (x, v)^\top \in \mathbb{R}^2$, output vector $\mathbf{y} = (x, v)^\top \in \mathbb{R}^2$, parameter vector $\mathbf{p} = (m, d)^\top \in \mathbb{R}^2$, and input force $u \in \mathbb{R}$. The nominal model parameters are set to $m = 1\,\text{kg}$ and $d = 20\,\text{Nsm}^{-1}$ for simulation experiments. The servo position is measured with an incremental encoder. The sampled measured position reads

$$x_{m;k} = x_k + e_k \tag{2.48}$$

where $e_k \sim \mathcal{N}\left(\mu, \sigma_x^2\right)$ is a Gaussian white noise process with $\mu = 0$ and $\sigma_x = 0.001\,\mathrm{m}$. Combining the model equations (2.47), its sensitivity equations, and the chosen metric of the Fisher matrix (D-criterion) with initial conditions, boundary conditions, and final time $t_e = 15\,\mathrm{s}$, we end up with the highly nonlinear optimal control problem

$$\begin{aligned}
&\max_u \det\left(\bar{\mathbf{F}}\left(u, \mathbf{p}\right)\right)\\
&\begin{pmatrix}\dot{x}\\ \dot{v}\end{pmatrix} = \begin{pmatrix}0 & 1\\ 0 & -\frac{d}{m}\end{pmatrix}\begin{pmatrix}x\\ v\end{pmatrix} + \begin{pmatrix}0\\ \frac{1}{m}\end{pmatrix} u\\
&\mathbf{x}\left(0\right) = \mathbf{0},\ u \in \left[-10\,\mathrm{N}, 10\,\mathrm{N}\right],\ x \in \left[0, 1\,\mathrm{m}\right],\ v \in \left[0.5\,\mathrm{m\,s^{-1}}, -0.5\,\mathrm{m\,s^{-1}}\right]\\
&\dot{\mathbf{S}} = \begin{pmatrix}0 & 1\\ 0 & -\frac{d}{m}\end{pmatrix}\mathbf{S} + \begin{pmatrix}0 & 0\\ \frac{1}{m^2}\left(dv - u\right) & -\frac{v}{m}\end{pmatrix},\ t > 0\\
&\mathbf{S}\left(0\right) = \mathbf{0}\\
&\mathbf{S}_y = \mathbf{S}\,.
\end{aligned} \tag{2.49}$$

**Basic Properties of the Normal Distribution**

**Proposition 2.1.**

(i) For $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ applies:

$$\mathrm{E}[X] = \mu\,, \qquad \mathrm{var}\left(X\right) = \sigma^2 \tag{2.50}$$

(ii) Let $X_1 \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ and $X_2 \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$ be two independent and normally distributed random variables. It applies:

$$X_1 + X_2 \sim \mathcal{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right) \tag{2.51}$$

(iii) For every $a, b \in \mathbb{R}$ applies

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \implies Y := aX + b \sim \mathcal{N}\left(a\mu + b, a\sigma^2\right)\,. \tag{2.52}$$

We assume both $x$ and $v$ are measured and used for the OID process. However, only the servo position is measured using an incremental encoder. The velocity is calculated from the sampled position through the backward difference quotient. With the standard noise assumption for the measured position $x_m = x + e$ from (2.9)

$$\begin{aligned} e_k &\sim \mathcal{N}\left(0, \sigma_x^2\right) \\ \mathrm{E}[e_k, e_{k+l}] &= \begin{cases} \sigma_x^2, & l = 0 \\ 0, & l \neq 0 \end{cases} \end{aligned} \tag{2.53}$$

and the basic properties of the normal distribution from Proposition 2.1, one obtains

$$e_{v;k} = \frac{e_k - e_{k-1}}{T} \sim \mathcal{N}\left(0, \frac{2}{T^2}\sigma_x^2\right) \tag{2.54}$$

for the noise characteristics of the approximated velocity. Taking into account (2.53), for the covariance applies

$$\begin{aligned} \operatorname{cov}\left(e_k, e_{v;k}\right) &= \mathrm{E}[\left(e_k - \bar{e}_k\right)\left(e_{v;k} - \bar{e}_{v,k}\right)] = \mathrm{E}[e_k e_{v;k}] \\ &= \mathrm{E}[e_k \left(\frac{e_k - e_{k-1}}{T}\right)] = \frac{1}{T}\,\mathrm{E}[e_k^2] - \frac{1}{T}\,\mathrm{E}[e_k e_{k-1}] \\ &= \frac{1}{T}\sigma_x^2 \,. \end{aligned} \tag{2.55}$$

Consequently, we get

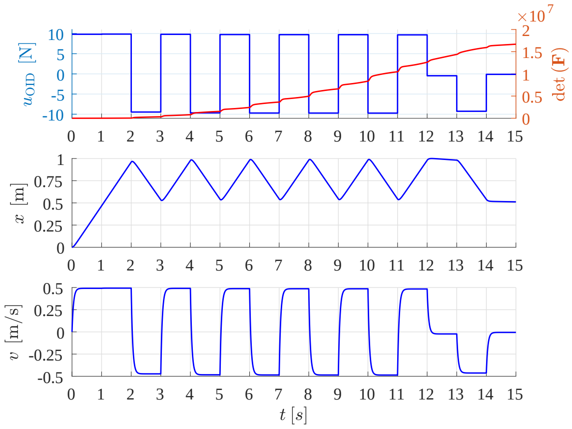
$$\left(e_k, e_{v;k}\right) \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \frac{1}{T}\sigma_x^2 \\ \frac{1}{T}\sigma_x^2 & \frac{2}{T^2}\sigma_x^2 \end{pmatrix}\right) \tag{2.56}$$

where

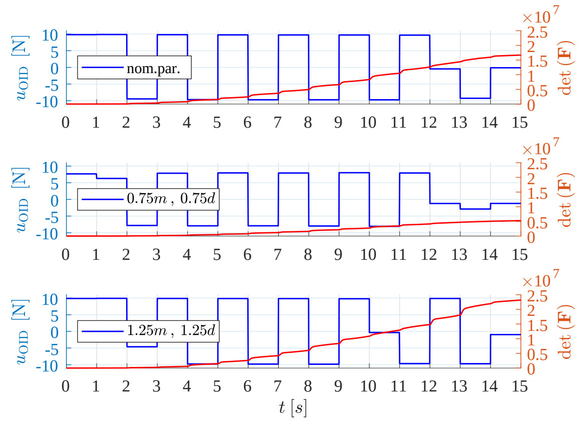
$$\boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_x^2 & \frac{1}{T}\sigma_x^2 \\ \frac{1}{T}\sigma_x^2 & \frac{2}{T^2}\sigma_x^2 \end{pmatrix}, \quad \boldsymbol{\sigma}^2 \in \mathbb{R}^{2\times 2} \tag{2.57}$$

is the covariance matrix, necessary for building the Fisher matrix. For reasons of comparability with an $n = 4$-bit PRBS signal, with maximum length sequence $2^n - 1 = 15$ and minimum hold time $\tau = 1\,\mathrm{s}$, the optimization algorithm also allows fifteen changes of the manipulated variable with the same minimum hold time. The OID result is shown in Fig. 2.7.

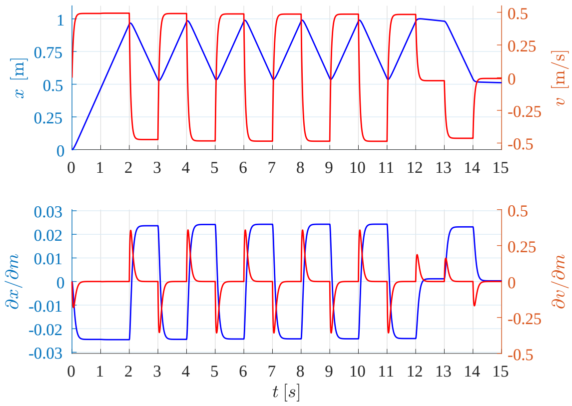
The optimal excitation signal looks very similar to a PRBS, which seems plausible because one knows from the literature, e.g., [2, 12, 35], that PRBS provides suitable excitation for linear systems. The slight deviations of the OID signal from possible PRBS signal characteristics are due to the compliance with boundary conditions and, above all, that the solution of the optimization task does not represent a global maxi-



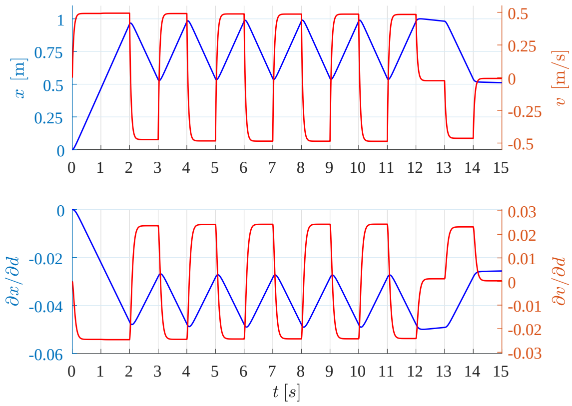

(a) Optimal input signal.




(b) Optimal input signal with uncertainties ($\pm 25\%$) in model parameters used for OID.




(c) Parameter sensitivities w.r.t vehicle mass $m$, assuming nominal model parameters.




(d) Parameter sensitivities w.r.t linear friction coefficient $d$, assuming nominal model parameters.

Figure 2.7: Optimal input design for the linear 1D-servo system using D-criterion.

mum. Fig. 2.7(b) shows the consequences of parameter uncertainties during the OID task. Despite different parameter values, the OID process results in similar signal characteristics. However, the parameter values' uncertainties may violate the initial conditions, final conditions, and state conditions. An iterative process of OID and parameter estimation may help here.

Due to its easy handling and low complexity, the OLS estimator is a proven tool to estimate unknown parameters of CT models. To avoid the direct calculation of the time derivatives of measurement signals, more sophisticated methods are available. One of them is called Poisson moment functionals (PMF) method, and one can interpret it as the convolution of a signal with the impulse response of a known stable filter. A known drawback of all methods for the exact elimination of time derivatives, e.g., PMF, is that the properties of measurement noise are changed. Thus, the least-squares estimator loses its pleasant property of unbiasedness or consistency, which is given under certain conditions. Chapter 3 provides more detailed information. Without going into details, the algebraic equation for identification reads

$$g_F^0(t) * u(t) = \begin{pmatrix} g_F^2(t) * x(t) & g_F^1(t) * x(t) \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix} \tag{2.58}$$

where

$$\begin{aligned} F^0(s) &= \frac{1}{(1+sT_f)^2}, & g_F^0(t) &= \mathcal{L}^{-1}\left\{F^0(s)\right\} \\ F^1(s) &= \frac{1}{(1+sT_f)^2}, & g_F^1(t) &= \mathcal{L}^{-1}\left\{F^1(s)\right\} \\ F^2(s) &= \frac{1}{(1+sT_f)^2}, & g_F^2(t) &= \mathcal{L}^{-1}\left\{F^2(s)\right\} \end{aligned} \tag{2.59}$$

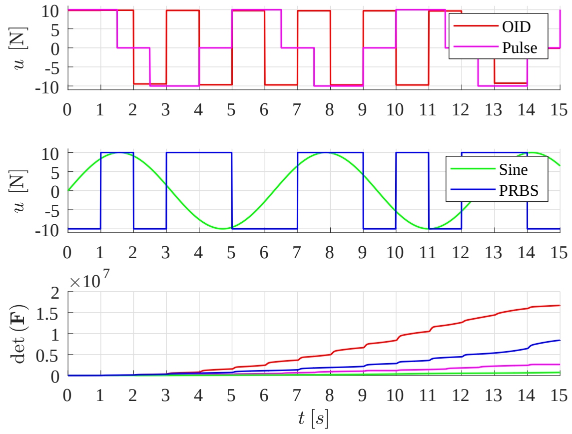
with filter time constant $T_f = 50\,\mathrm{ms}$. Evaluating (2.58) at $N > 2$ different points in time, the resulting set of equations is solved using OLS.

A simulation study with 1000 runs and a noisy system output $x_m(t) = x(t) + e(t)$, where $e \sim \mathcal{N}(0, 0.001^2)$, shows the effect of different input signals (OID, pulse pattern, sine excitation, and PRBS) for the parameter estimation result, see Tab. 2.2 or Fig. 2.8.

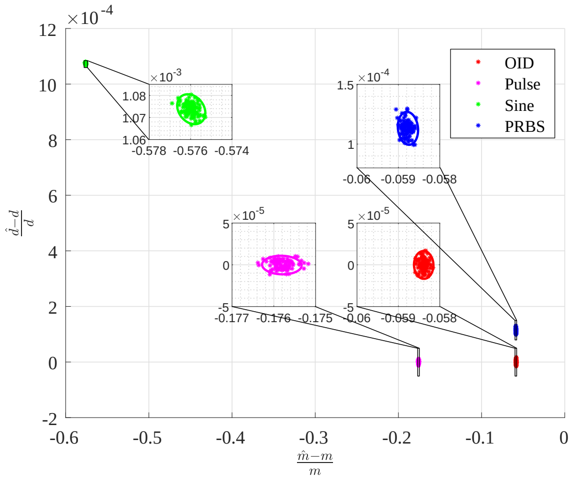
| nominal parameters | $m = 1\,\mathrm{kg}$ | | $d = 20\,\mathrm{Nsm}^{-1}$ | |
|---|---|---|---|---|
| input | $\hat{m}$ [kg] | | $\hat{d}$ [Nsm$^{-1}$] | |
| | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\mu}$ | $\hat{\sigma}$ |
| OID | 0.9416 | $9.89 \times 10^{-5}$ | 20 | $1.374 \times 10^{-4}$ |
| pulse-shaped | 0.8242 | $2.079 \times 10^{-4}$ | 20 | $9.11 \times 10^{-5}$ |
| sinusoidal | 0.4240 | $2.914 \times 10^{-4}$ | 20.0215 | $5.55 \times 10^{-5}$ |
| PRBS | 0.9412 | $9.89 \times 10^{-5}$ | 20.0023 | $1.128 \times 10^{-4}$ |

Table 2.2: Estimated parameter values $\hat{\mathbf{p}}$ are obtained by 1000 simulation runs with different input signals and noisy model output $x_{m;k} = x_k + e_k$ with additive Gaussian white noise, where $e_k \sim \mathcal{N}(\mu, \sigma_x^2)$ with $\mu = 0$ and $\sigma_x = 0.001\,\mathrm{m}$.

Obviously, the higher the information content of the excitation signal, the better the estimation result, i.e., the lower the bias. Especially for linear systems, the quality of the estimate thus correlates with the information content of the excitation signal. Interestingly, the PRBS signal does not provide the highest information content, although according to literature, it is an optimal excitation signal for linear systems, e.g., [11]. Please note that the PRBS sequence used here is one possible realization




(a) Different input signals with corresponding information content.

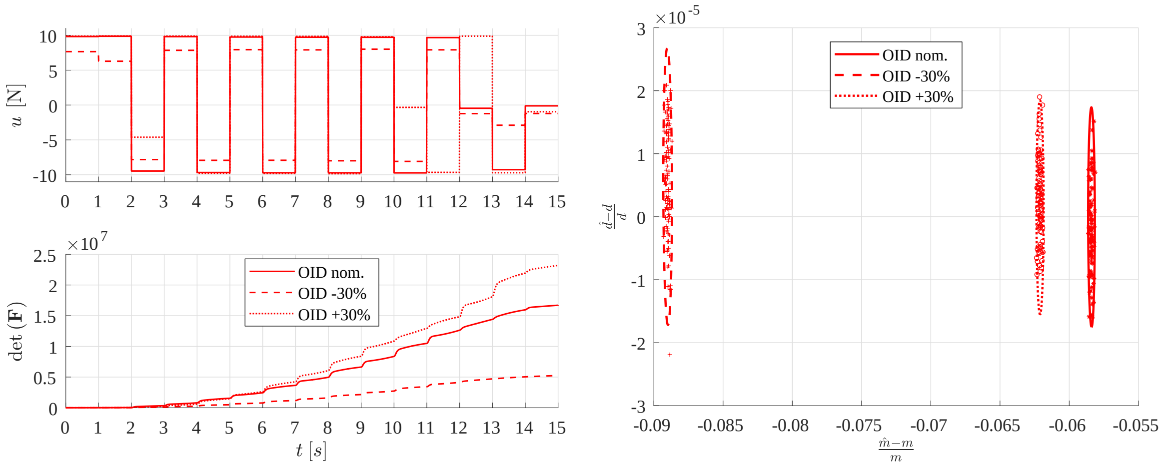



(b) Parameter covariance matrix of (normalized) estimation error with different input signals.

Figure 2.8: Effects of different excitation signals (OID signal, pulse pattern, sine wave and PRBS) with its different information content on the parameter estimation for the linear 1D-servo system. For a better readability of the diagram only every tenth data point is shown.

of a four-bit PRBS sequence. Depending on the actual realization, the information content varies. Hence, other PRBS realizations may provide a higher information content than the OID signal. The estimation result shows that the PRBS signal provides almost the same good results as the OID signal. In other words, the OID signal gives equally good or even slightly better results than the PRBS signal, which provides optimum excitation for linear systems. Interestingly, the friction coefficient $d$ is estimated relatively accurately by the usage of all signals. However, there are large differences in the estimation of the mass $m$. The reason for this is the different weak or strong acceleration phases. In general, the higher the car acceleration, the more accurately the mass is estimated.

The effects of parameter uncertainties during the generation of the excitation signals on their information content and the parameter estimation are shown in Fig. 2.9. Tendentially, the higher the information content, the more accurate the estimation result. The arbitrary scalar quantity, which is used as a norm for the parameter covariance matrix, inseparably blends the parts of the different parameters. Thus, one cannot draw any conclusions about the accuracy of estimates of individual parameters from the absolute value of the information content. However, Fig. 2.9 shows that the scalar quality functionals derived from the parameter covariance matrix are not always completely accurate. Despite the slightly higher information content of the red-dotted colored signal compared to the red-solid one, it is exactly the way the other round is for the estimation error. Nevertheless, a tendency can be derived. The significantly lower information content of the red-dashed colored signal results in the most inaccurate estimation result. As mentioned above, the model parameters necessary for the OID




(a) Different input signals with corresponding information content.

(b) Parameter covariance matrix of (normalized) estimation error with different input signals.

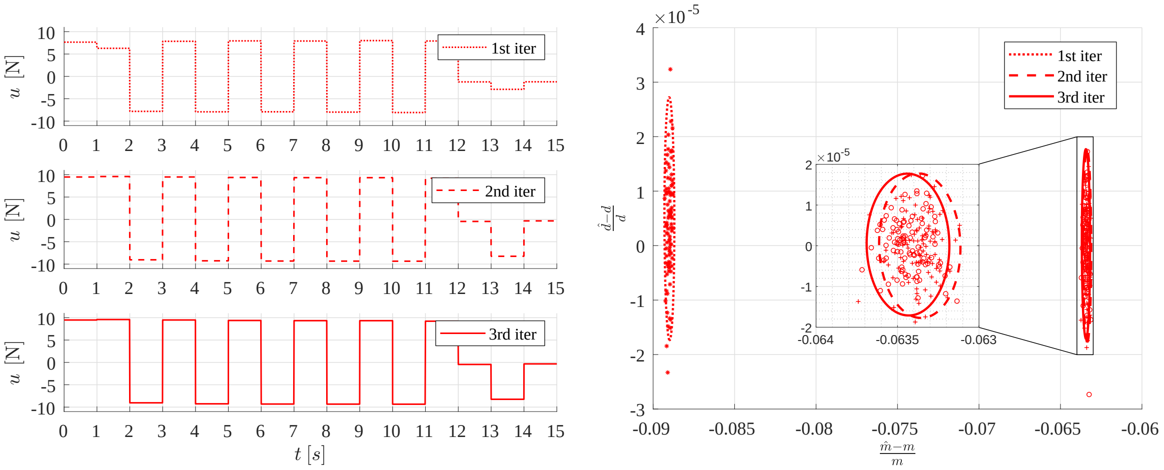
Figure 2.9: Effects of parameter uncertainties during OID task on the parameter estimation for the linear 1D-servo system.[5]

task are subject to uncertainties. Hence, an iterative process of OID and parameter estimation is recommended. For demonstration purpose, the OID task is initiated with parameter values that differ by 30% from the nominal parameters. With the excitation signal obtained from this initial OID task, in a first iteration, the parameters are estimated, and the OID process is rerun through to obtain a more suitable input signal. This process is repeated until it converges. Figure 2.10 and Tab. 2.3 show that the iterative OID process for this application already converges with the second iteration.

[5]Because OID-generated signals are shown in red in all other figures, different OID signals are not distinguished by colors but by line styles.

| iteration | $\hat{m}$ [kg] | | $\hat{d}$ [Nsm$^{-1}$] | |
|---|---|---|---|---|
| | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\mu}$ | $\hat{\sigma}$ |
| 0 | 0.75 | - | 15 | - |
| 1 | 0.9110 | $1.281 \times 10^{-4}$ | 20.0001 | $1.914 \times 10^{-4}$ |
| 2 | 0.9366 | $1.013 \times 10^{-4}$ | 20 | $1.463 \times 10^{-4}$ |
| 3 | 0.9366 | $1.014 \times 10^{-4}$ | 20 | $1.448 \times 10^{-4}$ |

Table 2.3: Iterative process of OID and parameter estimation. Initial guess for OID task: $\hat{m} = 0.75\,\text{kg}$ and $\hat{d} = 15\,\text{N}\,\text{m}\,\text{s}^{-1}$. Nominal values: $m = 1\,\text{kg}$ and $d = 20\,\text{N}\,\text{m}\,\text{s}^{-1}$. Estimated parameter values $\hat{\mathbf{p}}$ are obtained by 1000 simulation runs and noisy model output $x_{m;k} = x_k + e_k$ with additive Gaussian white noise, where $e_k \sim \mathcal{N}(\mu, \sigma_x^2)$ with $\mu = 0$ and $\sigma_x = 0.001\,\text{m}$.




(a) Iteratively obtained optimal input signals.

(b) Parameter covariance matrix of (normalized) estimation error for iterative OID task.

Figure 2.10: Iterative process of OID and parameter estimation for the linear 1D-servo system. Initial guess for OID task: $\hat{m} = 0.75\,\text{kg}$ and $\hat{d} = 15\,\text{N}\,\text{m}\,\text{s}^{-1}$. Thousand simulation runs are performed. Output noise is $x_{m;k} = x_k + e_k$ with $e_k \sim \mathcal{N}(\mu,\, \sigma_x^2)$, where $\mu = 0$ and $\sigma_x = 0.001\,\text{m}$.

## 2.6.2 Nonlinear Model (3 Parameters)

The extension of the linear friction model from Section 2.6.1 by a static component, i.e., Coulomb friction, results in the nonlinear 1D-servo model with its equations

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \\ -\frac{d}{m} v - \frac{F_c}{m} \operatorname{sgn}(v) + \frac{1}{m} u \end{pmatrix} \tag{2.60}$$

with state vector $\mathbf{x} = (x,\, v)^\top \in \mathbb{R}^2$, output vector $\mathbf{y} = (x, v)^\top \in \mathbb{R}^2$, parameter vector $\mathbf{p} = (m, d, F_\mathrm{c})^\top \in \mathbb{R}^3$, and input $u \in \mathbb{R}$. The nominal model parameters are set to $m = 1\,\mathrm{kg}$, $d = 20\,\mathrm{Nsm}^{-1}$ and $F_\mathrm{c} = 2\,\mathrm{N}$ for simulation experiments. Again, the servo position is measured using an incremental encoder. For the sampled measured position applies

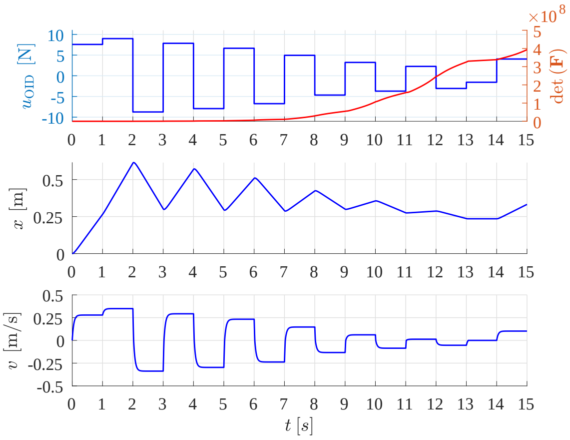
$$x_{m;k} = x_k + e_k \tag{2.61}$$

where $e_k \sim \mathcal{N}\left(\mu, \sigma_x^2\right)$ with $\mu = 0$ and $\sigma_x = 0.001\,\mathrm{m}$. Combining the model equations (2.60), its sensitivity equations according to (2.33), and the chosen metric of the Fisher matrix (D-criterion) with initial conditions and state constraints, we end up with a highly nonlinear optimal control problem

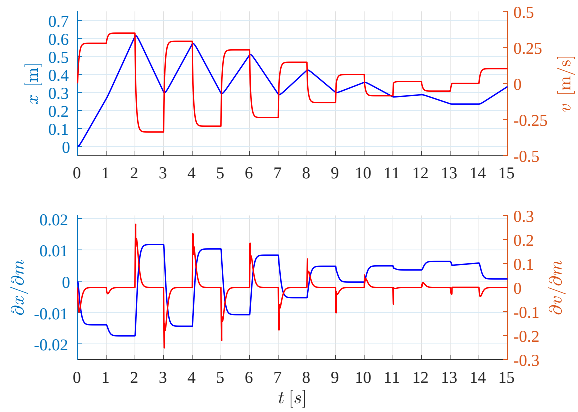
$$\begin{aligned}
&\max_u \det\left(\bar{\mathbf{F}}\left(u, \mathbf{p}\right)\right)\\
&\begin{pmatrix}\dot{x}\\ \dot{v}\end{pmatrix} = \begin{pmatrix} v \\ -\frac{d}{m}v - \frac{F_\mathrm{c}}{m}\operatorname{sgn}\left(v\right) + \frac{1}{m}u\end{pmatrix}\\
&\mathbf{x}\left(0\right) = \mathbf{0},\, u \in \left[-10\,\mathrm{N}, 10\,\mathrm{N}\right],\, x \in \left[0, 1\,\mathrm{m}\right],\, v \in \left[0.5\,\mathrm{m\,s^{-1}}, -0.5\,\mathrm{m\,s^{-1}}\right]\\
&\dot{\mathbf{S}} = \begin{pmatrix} 0 & 1 \\ 0 & -\frac{1}{m}\left(d + F_\mathrm{c}\frac{\mathrm{d}}{\mathrm{d}t}\left(\operatorname{sgn}\left(v\right)\right)\right)\end{pmatrix}\mathbf{S}\\
&\quad + \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{m^2}\left(dv + F_\mathrm{c}\operatorname{sgn}\left(v\right) - u\right) & -\frac{v}{m} & -\frac{\operatorname{sgn}(v)}{m}\end{pmatrix},\; t > 0\\
&\mathbf{S}\left(0\right) = 0\\
&\mathbf{S}_y = \mathbf{S}
\end{aligned} \tag{2.62}$$

with final time $t_e = 15\,\mathrm{s}$. Please note that the approximation $\operatorname{sgn}\left(v\right) \approx \tanh\left(kv\right), k = 1 \times 10^3$ is introduced to be able to calculate $\frac{\mathrm{d}}{\mathrm{d}t}\left(\operatorname{sgn}\left(v\right)\right)$ numerically. The error covariance matrix, which is necessary to build the Fisher matrix, is again calculated using (2.57). Figure 2.11 shows, that in contrast to the linear case, the generated optimal excitation signal is not a signal similar to PRBS but consists of different amplitudes. Different amplitudes seem reasonable because the estimation algorithm must distinguish between linear friction and static friction. The OID task automatically introduces this additional system knowledge when generating the excitation signal. For the estimation of the unknown model parameters, the PMF approach combined with OLS is used again.

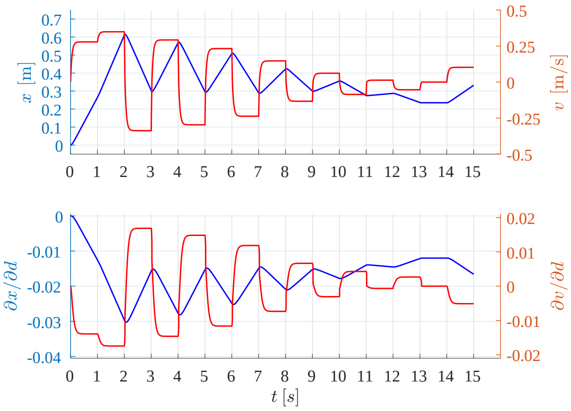
Analogous to the linear model (2.58), one obtains the identification equation, linear



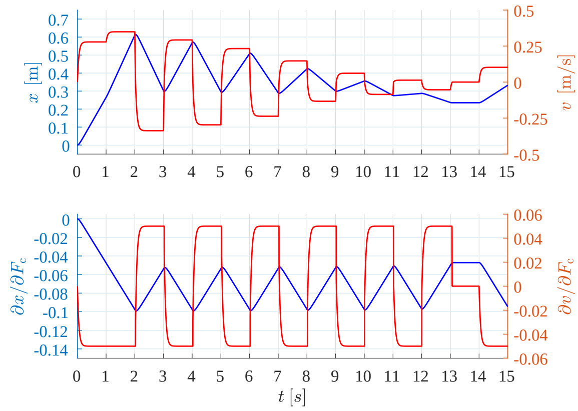

(a) Optimal input signal.




(b) Parameter sensitivities w.r.t vehicle mass $m$, assuming nominal model parameters.




(c) Parameter sensitivities w.r.t linear friction coefficient $d$, assuming nominal model parameters.




(d) Parameter sensitivities w.r.t Coulomb friction $F_c$, assuming nominal model parameters.

Figure 2.11: Optimal input design for nonlinear 1D-servo system using D-criterion.

in the parameters,

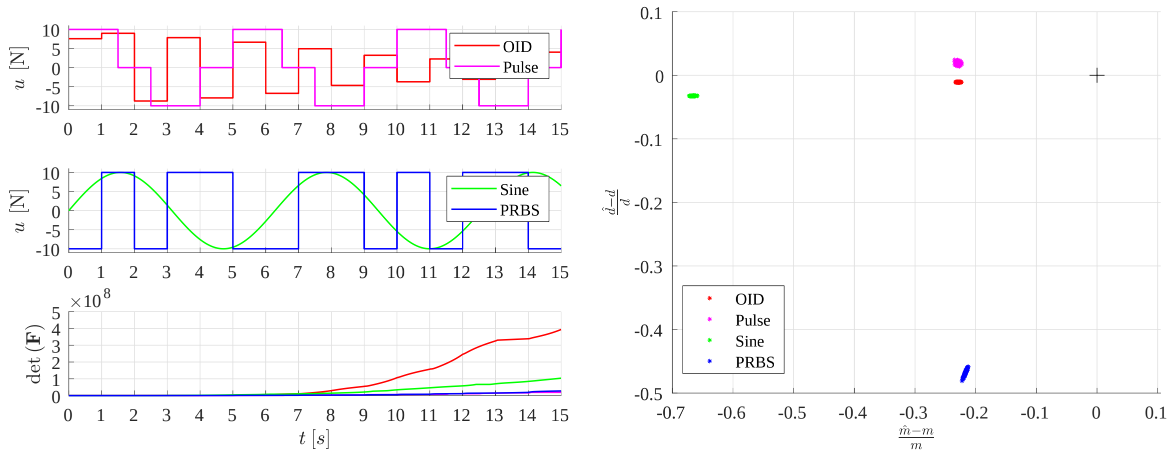
$$g_F^0(t) * u(t) = \begin{pmatrix} g_F^2(t) * x(t) & g_F^1(t) * x(t) & g_F^0(t) * \mathrm{sgn}(v(t)) \end{pmatrix} \begin{pmatrix} m \\ d \\ F_c \end{pmatrix} \tag{2.63}$$

where

$$\begin{aligned} F^0(s) &= \frac{1}{(1+sT_f)^2}, & g_F^0(t) &= \mathcal{L}^{-1}\left\{F^0(s)\right\} \\ F^1(s) &= \frac{s}{(1+sT_f)^2}, & g_F^1(t) &= \mathcal{L}^{-1}\left\{F^1(s)\right\} \\ F^2(s) &= \frac{s^2}{(1+sT_f)^2}, & g_F^2(t) &= \mathcal{L}^{-1}\left\{F^2(s)\right\} \end{aligned} \tag{2.64}$$

with filter time constant $T_f = 50\,\text{ms}$. Again, evaluating at $N > 2$ different points in time, the resulting set of equations is solved using OLS.

A simulation study with 1000 runs and a noisy system output $x_m(t) = x(t) + e(t)$, where $e \sim \mathcal{N}(0, \sigma_x^2)$ with $\sigma_x = 0.001\,\text{m}$, shows the effect of different input signals (OID, pulse pattern, sine excitation, and PRBS) for the parameter estimation result, see Fig. 2.12 and Fig. 2.13, respectively.




(a) Different input signals with corresponding information content.

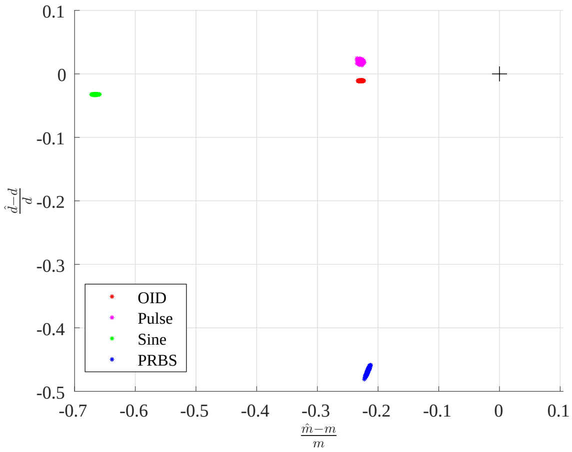
(b) Parameter covariance matrix of (normalized) estimation error with different input signals.

Figure 2.12: Effects of different excitation signals (OID signal, pulse pattern, sine wave and PRBS) with its different information content on the parameter estimation for the nonlinear 1D-servo system. For a better readability of the diagram only every tenth data point is shown.

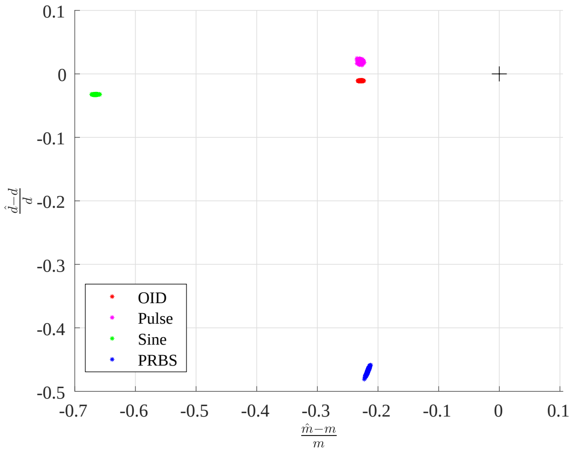
Table 2.4 numerically presents the graphical results from Fig. 2.12 and Fig. 2.13. The OID signal contains the highest information content and provides the parameter estimates with the lowest bias, i.e., the most accurate estimates. Although the single

| nominal parameters | $m = 1\,\text{kg}$ | | $d = 20\,\text{Nsm}^{-1}$ | | $F_c = 2\,\text{N}$ | |
|---|---|---|---|---|---|---|
| input | $\hat{m}$ [kg] | | $\hat{d}$ [Nsm$^{-1}$] | | $\hat{F}_c$ [N] | |
| | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\mu}$ | $\hat{\sigma}$ |
| OID | 0.7716 | 0.0023 | 19.7844 | 0.0060 | 2.0531 | 0.0014 |
| pulse-shaped | 0.7707 | 0.0025 | 20.3888 | 0.0431 | 1.8495 | 0.0166 |
| sinusoidal | 0.3339 | 0.0026 | 19.3495 | 0.0048 | 2.2164 | 0.0015 |
| PRBS | 0.7828 | 0.0022 | 10.6175 | 0.0989 | 5.7073 | 0.0391 |

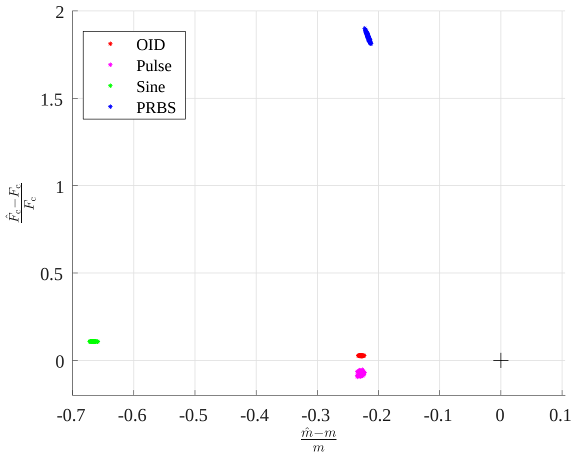
Table 2.4: Estimated parameter values $\hat{\mathbf{p}}$ obtained by 1000 simulation runs with different input signals and noisy model output $x_{m;k} = x_k + e_k$ with additive gaussian white noise, where $e_k \sim \mathcal{N}(\mu, \sigma_x^2)$ with $\mu = 0$ and $\sigma_x = 0.001\,\text{m}$.



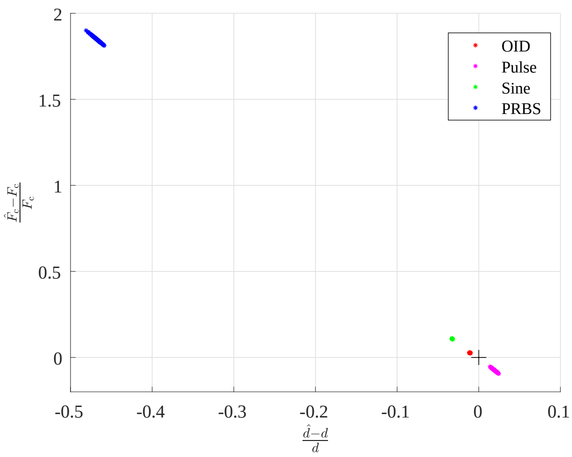

(a) Parameter covariance matrix of (normalized) estimation error with different input signals.




(b) Sectional view of the parameter covariance matrix in the $md$ plane.




(c) Sectional view of the parameter covariance matrix in the $mF_c$ plane.




(d) Sectional view of the parameter covariance matrix in the $dF_c$ plane.

Figure 2.13: Sectional views of the parameter covariance matrix of the (normalized) estimation error with different excitation signals (OID signal, pulse pattern, sine wave and PRBS) for the nonlinear 1D-servo system. For a better readability of the diagram only every tenth data point is shown.

sine wave and the PRBS signal also have a high information content, the respective estimation results are strongly biased. For example, the bang-bang behavior of PRBS signals prevents distinguishing between linear and static frictional components, which requires phases with different speeds to distinguish between static and velocity proportional friction. The user provides this additional system knowledge, e.g., for the pulse-shaped signal, while the OID signal includes such phases automatically due to the OID process. To sum up, the generation of optimal input signals with high information content leads to more accurate parameter estimates, i.e., reducing the bias.

Conversely, not every excitation signal with high information content is suitable for each system, especially for nonlinear systems.

## 2.7 Conclusion

Using a simple first-order system, one can show that the optimal selection of the evaluation time points using DoE, i.e., with the help of the Fisher matrix or quantities derived from it, significantly improves the parameter estimation result. In practice, the broad topic of DoE is often reduced to the excitation signal's appropriate choice (OID), which is why this work focuses on this particular issue of DoE. Using the practice-relevant example 1D-servo system, one can show the improvements for the parameter estimation. The optimal excitation signal for linear systems is similar to the PRBS signal (see Fig. 2.7(a)). This result supports that the coverage of the "entire" frequency spectrum is crucial. For nonlinear systems, the amplitude level also plays a role in the quality of excitation signals. The signals have to be adapted to the task at hand, i.e., to the model. For example, the distinction between Coulomb static friction and linear friction requires velocity phases with different amplitudes. For this reason, the PRBS signal gives poor results of the friction terms for the nonlinear servo model, see Fig 2.13 or Tab. 2.4. The OID process introduces this additional model knowledge without user intervention.

An additional benefit of OID is the generation of signals, which, assuming sufficiently accurate initial values of the model parameters, guarantee compliance with specified bounds on the state variables, their initial and final values, and requirements on the input signal. One recommends an iterative OID and parameter estimation process for significant parameter uncertainties until the OID result converges (see Fig. 2.10).

# 3 Parameter Estimation with Focus on Parametric Continuous-Time Systems

The topic of system identification is essential for most applications in engineering, especially for automatic control. As already mentioned in the introduction, [1] divides system identification into four major issues: experiment data generation, choice of model structure, estimation of unknown model parameters, and verification of identification results. In [36], an object referred to as a system is known through modeling and identification, i.e., that modeling and identification methods help people obtain knowledge about systems.

The modeling topic is a comprehensive one and includes various and well-established methods, e.g., physical modeling. Physical (dynamical) systems are usually modeled by applying physical laws and phenomena, especially in control engineering. Typically, these systems are native in the continuous-time (CT) domain and are described by parametric models using differential equations. Thus, the model structure and the key parameters are a result of the physical modeling process. Numerical values of unknown model parameters are obtained by parameter estimation. For these reasons, in this thesis, we focus on the parameter estimation approaches for parametric models.

## 3.1 State of the Art

Of course, the topic parameter estimation is not a new one. Numerous different methods for estimating unknown model parameters have been developed and are well-established. Over the years, two significant directions have evolved in the identification of continuous-time (CT) systems, namely the identification of discrete-time (DT) models and transform back into the CT domain (indirect approach) and the straightforward identification of CT models (direct approach). For example, in [36–38], a detailed survey of identification methods focusing on parametric CT models can be found. In the 1970s and 1980s, methods dealing with DT models were dominating, partly due to the development of digital computers at that time, see [10, 11, 39]. In parallel, the development of CT approaches was initiated and published in 1981 [40], followed by [37, 38], to name but a few.

Although this thesis focuses only on identifying CT models, we briefly summarize selected basic methods in this section for DT models. Besides the completeness of the literature review, this is mainly since most of the investigations and methods are derived for DT models first and then (if possible) applied or extended to CT models. Besides, we restrict ourselves from identifying parametric linear single-input single-output (SISO) models in the time domain or identifying models which are at least linear in the parameters. Identification methods and approaches for multiple-input multiple-output (MIMO) models are treated in, e.g., [2, 3], and the identification of nonlinear models [2, 3, 11]. Identification in the frequency domain and identification of non-parametric models are discussed in, e.g., [2, 3, 10, 11].

### 3.1.1 Identification of Linear SISO Parametric DT Systems

Although most technical processes are CT processes, DT mathematical models are often derived. The main reasons for this are the less complicated mathematical handling in parameter estimation and the necessary digital implementation [3].

#### 3.1.1.1 Error Models

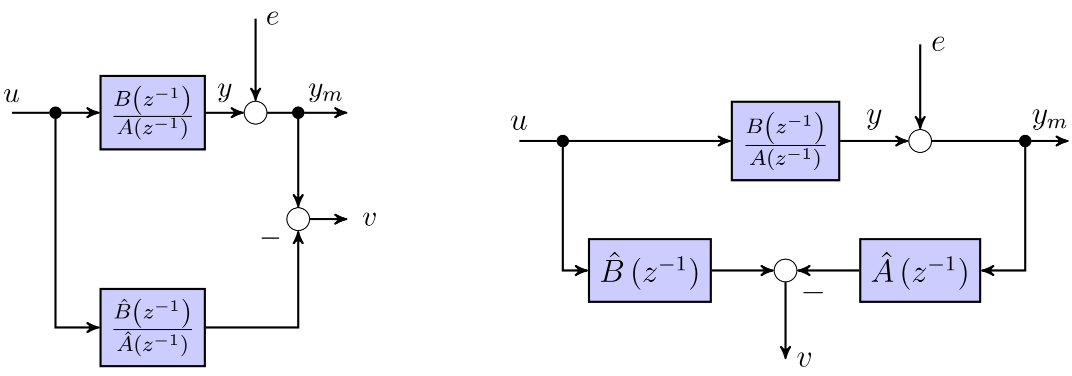
The deviation between the process (system) and the corresponding model describes the quality of a model or estimated model parameters. Different interpretations are possible, e.g., [39]. One can distinguish between

(i) input error,

(ii) output error (OE), and

(iii) (generalized) equation error (EE),

where only output error (OE) and equation error (EE) are suitable for open-loop system identification. Please note that equation error methods (EEMs) are often indicated as prediction error methods (PEMs). In principle, both output error methods (OEMs) and EEMs, lead to correct parameter estimates. However, for practical reasons, different error measures are more suitable for certain types of models than others,




(a) Output error. (b) Generalized equation error.

Figure 3.1: Different error measures (error models) between system and model.

e.g., [2, 41, 42]. Hence, the error model used depends on the problem at hand. Exemplarily, it is helpful to obtain linearity in the parameters to be identified from a mathematical perspective. Therefore, one chooses the output error for non-parametric models, e.g., impulse responses, while selecting the (generalized) equation error for parametric models, i.e., differential equations, difference equations, and transfer functions.

In contrast to the output error, it is evident that the noise properties changes using the generalized equation error, see Fig. 3.1. This yields to the issue of the bias of estimators tackled later.

#### 3.1.1.2 Least-Squares Method

Consider the DT transfer function

$$G\left(z^{-1}\right) = \frac{y\left(z\right)}{u\left(z\right)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_{m_d} z^{-m_d}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}} \tag{3.1}$$

with time-domain representation

$$y_k + a_1 y_{k-1} + \cdots + a_n y_{k-n} = b_0 u_k + b_1 u_{k-1} + \cdots + b_n u_{k-n} \tag{3.2}$$

where, e.g., $y\left(t_k\right) = y_k$. Replacing the model output $y$ by measured (disturbed) values

$$y_{m;k} = y_k + e_k \tag{3.3}$$

where $e_k$ is a stochastic noise, one obtains

$$\underbrace{y_{m;k}}_{y_{LSQ;k}} = \underbrace{\begin{pmatrix} -y_{m;k-1} & \dots & -y_{m;k-n} & u_k & \dots & u_{k-n} \end{pmatrix}}_{=\mathbf{w}_k^\top} \underbrace{\begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b_0 \\ \vdots \\ b_n \end{pmatrix}}_{=\mathbf{p}} + v_k \,. \tag{3.4}$$

where $v_k$ is called generalized equation error (EE). Evaluating (3.4) at $N > n_p$ , $n_p = 2n + 1$ discrete samples gives

$$\mathbf{y}_{LSQ} = \mathbf{W}\mathbf{p} + \mathbf{v} \tag{3.5}$$

where $\mathbf{y}_{LSQ} \in \mathbb{R}^N$, $\mathbf{W} \in \mathbb{R}^{N\times n_p}$, $\mathbf{p} \in \mathbb{R}^{n_p}$, and $\mathbf{v} \in \mathbb{R}^N$. The optimal solution in the sense of least-squares reads

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top\mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \,. \tag{3.6}$$

For convergence studies, (3.5) is inserted in (3.6). The expected value reads

$$\mathrm{E}[\hat{\mathbf{p}}] = \mathbf{p} + \mathrm{E}[\left(\mathbf{W}^\top\mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{v}] = \mathbf{p} + \Delta\mathbf{p} \tag{3.7}$$

where $\Delta\mathbf{p} = \mathrm{E}[\left(\mathbf{W}^\top\mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{v}]$ is a bias.

A very well-known but artificially generated configuration exists for which the least-squares estimator delivers unbiased results, c.f. Fig. 3.2. Literature, e.g., [2], shows that if $e_k$ is generated by filtering white noise $n_k$ with the plant's denominator, i.e.,

$$e\left(z\right) = \frac{1}{A\left(z^{-1}\right)} n\left(z\right) \tag{3.8}$$

the bias becomes zero, even in finite time. Apart from the artificially created setting, i.e., for practical applications, the estimator is biased for discrete-time models but at least consistent or asymptotically unbiased, e.g., [3, 43].

For the least-squares algorithm, some extensions and adaptations exist. For example, the recursive least-squares (RLS) method, introduced by [39, 44, 45], estimates the unknown system parameters in real time. Weighted least-squares (WLS) is another extension, where the different entries of the equation error are weighted differently. If the covariance matrix of the equation error is known, and one chooses the weighting

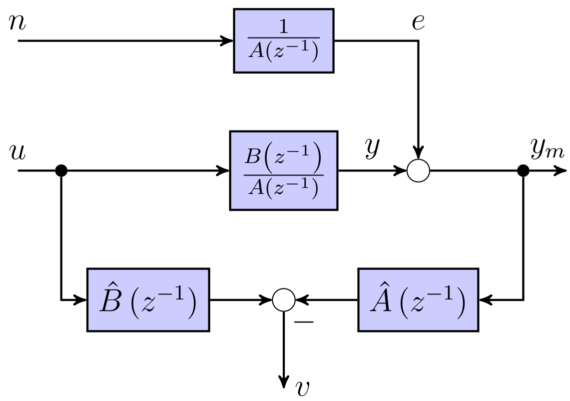



Figure 3.2: Block diagram of a required structure for an unbiased estimation of DT transfer functions, where $n \sim i.i.d\,(0, \sigma^2)$, e.g., Gaussian white noise.

matrix $\mathbf{Q}$ as

$$\mathbf{Q} = \left(\mathrm{E}[\mathbf{v}\mathbf{v}^\top]\right)^{-1} \tag{3.9}$$

one obtains parameter estimates

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top\mathbf{Q}\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{Q}\mathbf{y}_{LSQ}\,. \tag{3.10}$$

with minimum variance [46,47]. For the weighted least-squares (WLS) algorithm, there is also a recursive form for online parameter estimation. Weighting newer measurement data more heavily than older ones enables the algorithm to estimate slowly time-varying parameters. This unique choice of weighting is also known as exponential forgetting.

In literature, there exist several least-squares-based methods to avoid biased estimation results. The basic idea is to introduce special noise form filters to transform the initially correlated equation error into an uncorrelated one [2]. This approach includes the generalized least-squares (GLS) algorithm, e.g., [48–50] with its recursive implementation [51], as well as the extended least-squares (ELS) algorithm [52, 53]. The method of total least-squares (TLS), also known as *errors-in-variables*, assumes disturbances or error in the observation vector $\mathbf{y}_{LSQ}$ and the data matrix $\mathbf{W}$. This approach is related to the principal component analysis (PCA) find correlations in data sets and reduce dimensionality. For a detailed survey, the kindly reader is referred to, e.g., [54–56]. For example, an iterative TLS algorithm is presented in [57].

Another possibility to obtain at least asymptotically unbiased least-squares results is to estimate and correct the bias itself. These approaches are summarized under the collective term *bias compensation methods*. Such methods estimate the error variance(s) and deal with discrete-time noisy output and noisy input-output models, e.g., [58–61].

#### 3.1.1.3 Instrumental Variables Method

The instrumental variables (IV) method, e.g., [2, 3, 43, 62–65], is another modification of the least-squares algorithm to obtain asymptotically unbiased parameter estimation results. As stated in [64], the least-squares estimation's asymptotic bias can be eliminated by introducing special instrumental variables (IV) and correlation analysis. These IVs are chosen to be highly correlated with input and output signals but stochastically independent of the measurement noise. As instrumental variables, one can choose the input signal [66] or even stronger correlated, an estimate of the undisturbed model output obtained by the known input signal and the current parameter estimates, e.g., [67]. Selecting suitable instrumental variables is crucial for the parameter estimation quality and thus the crucial aspect of this method. If the correlation between IVs and undisturbed signals is insufficient, the parameter estimates are consistent but show a considerable variance [64].

#### 3.1.1.4 Bayes and Maximum Likelihood Methods

If one considers the parameters from a statistical perspective, they are random variables (RVs), e.g., [2, 43]. In Bayes estimation, the parameters are (statistically) described by its probability density function (PDF) $f_{\mathrm{p}}(p)$ and the model outputs by the conditional probability density function (PDF) $f_{\mathrm{y}}(y|p)$. Since the parameters' distribution function is usually unknown in practice, the Bayesian approach is of theoretical interest. However, the Bayes method forms the basis of other estimators, such as the Maximum Likelihood (ML) estimation or the ordinary least-squares (OLS) estimator [2, 3]. If no information about the distribution density function is available and assumed to be equally distributed, the Bayes estimator simplifies to the Maximum Likelihood (ML) estimator, described in, e.g., [10, 11, 68]. Assuming independent and identically distributed (i.i.d.) and Gaussian distributed noise, it can be shown that the ML approach leads to asymptotically efficient parameter estimates, i.e., reaches the Cramér-Rao lower bound (CRLB), [3]. If the noise assumptions are further restricted, i.e., i.i.d. normally distributed noise, the ML estimator coincides with the OLS estimator. In [69], a recursive algorithm of the ML approach for linear DT models is introduced.

### 3.1.2 Identification of Linear SISO Parametric CT Systems

Although methods for direct identification of CT systems, i.e., CT models, have been relegated to the background for some time due to the continuing trend towards digitization, they have some advantages over DT methods that should not be disregarded. An overview of existing methods and approaches for identifying CT models is given in [70]. The essential advantages are worked out in [36] and briefly summarized here:

(i) CT models or their parameters provide a good insight into the system properties.

(ii) CT models preserve knowledge about (parts of) the system: The discretization process is associated with some unwanted effects. For example, the discretization of a strictly proper CT rational transfer function with $n$ poles results in a rational DT transfer function with $n-1$ zeros. The zeros can not be represented in closed form as functions of the CT system parameters and the sampling time, with the direct consequence that the discretized model's parameters have no direct relation to the physical model parameters, e.g., [71]. For example, $G(s) = \frac{d}{s^3+as^2+bs+c}$ becomes $G(z) = \frac{b_1+b_2z+b_3z^2}{z^3+a_1z^2+a_2z+a_3}$, i.e., three numerator parameters have to be estimated instead of one. Hence, existing prior knowledge of CT models is wholly lost through discretization.

(iii) Discretization can lead to unwanted problems at high sampling rates, see [36]. The results of conventional DT methods ($z = e^{sT}$) do not converge to CT-based methods when the sampling time tends to zero, i.e., $T \to 0$. There exist unconventional DT methods for discretizing CT models, avoiding problems at high sampling rates, see, e.g., [72, 73].

(iv) Discretization may transform a causal and stable CT transfer function into a non-minimum phase DT transfer function. For more detailed information, the kindly reader is referred to [36, 74].

#### 3.1.2.1 Determination of the Time Derivatives

In contrast to DT methods, CT methods provide an extra challenge, namely handling time derivatives. In any case, the aim is the transformation of a differential equation into an algebraic equation. If the time derivatives are not measurable, they must be determined from the sampled input signals $u(t_k)$ and output signals $y(t_k)$. There are two different approaches, the approximation by numerical differentiation and the exact transformation of the time derivatives, i.e., avoid explicitly computing the time derivatives of sampled signals. As mentioned in [3,36], especially the following methods are suitable for this purpose:

(i) Modulating function method (MFM)

(ii) Poisson moment functionals (PMF)

(iii) Integral equation approach

In any case, it should be noted that any method for avoiding the explicit calculation of the derivatives are derived assuming non-random (deterministic) input and output signals. In practice, at least the output signals are noise affected, which changes the noise properties, e.g., [11]. From this, it follows directly that the parameter estimation results are generally biased. This thesis focuses on determining the estimation bias caused by such transformations for the class of PMFs in detail and MFs in principle.

#### 3.1.2.1.1 Numerical Differentiation

Numerical differentiation is perhaps the most obvious way to determine the unknown time derivatives. Possible variants are the backward difference quotient

$$\hat{\dot{y}}\left(t_k\right) = \frac{y\left(t_k\right) - y\left(t_{k-1}\right)}{T} \tag{3.11}$$

the forward difference quotient

$$\hat{\dot{y}}\left(t_k\right) = \frac{y\left(t_{k+1}\right) - y\left(t_k\right)}{T} \tag{3.12}$$

and the central differential quotient

$$\hat{\dot{y}}\left(t_k\right) = \frac{y\left(t_{k+1}\right) - y\left(t_{k-1}\right)}{2T} \tag{3.13}$$

where $T$ denotes the sample time, and the accent *hat* indicates the approximative character. Usually, measurement signals are noise-affected, so numerical differentiation leads to unsatisfactory results in practical application, especially for higher time derivatives. For better results, e.g., [75], uses interpolating functions.

#### 3.1.2.1.2 Modulating Functions (MF) Approach

As stated in [36,76], the modulating function method (MFM) was introduced by [77] in 1957. The MFM's central idea is to shift the time derivation from a measured variable to an adequately defined modulating function (MF), achieved by partial integration.

**Modulating Function**

**Definition 3.1.** A function $\varphi_k(t) : [0, T'] \to \mathbb{R}$ is called modulating function (MF) of adequate order $k$ if

$$\varphi_k^{(i)}(0) = \varphi^{(i)}(T') = 0 \tag{3.14}$$

for $i = 0, 1, 2, \dots, k-1$.

Considering integration by parts and (3.14), multiplying a signal $y$ by $\varphi_k$ and integration leads to the fundamental equation of the MFM

$$\int_0^{T'} \varphi_k(t)\, y^{(i)}(t)\, \mathrm{d}t = \int_0^{T'} (-1)^i \varphi_k^{(i)}(t)\, y(t)\, \mathrm{d}t\,. \tag{3.15}$$

This approach's smart property is that the initial and final values of the signals vanish due to Definition 3.1.

Let us consider a simple first-order model with its differential equation

$$y(t) = -T_1 \dot{y}(t) + V u(t) \tag{3.16}$$

with output $y(t) \in \mathbb{R}$, input $u(t) \in \mathbb{R}$, and parameter vector $\mathbf{p} = (T_1, V)^\top \in \mathbb{R}^{n_p}$. Furthermore, let us assume that the input-output data is available within the time interval $[0, T']$. Multiply the differential equation by $\varphi_k(t)$ and integrate over $[0, T']$, and taking into account (3.15), gives

$$\int_0^{T'} \varphi_k(t)\, y(t)\, \mathrm{d}t = -T_1 \int_0^{T'} \varphi_k(t)\, \dot{y}(t)\, \mathrm{d}t + V \int_0^{T'} \varphi_k(t)\, u(t)\, \mathrm{d}t\,. \tag{3.17}$$

Using (3.15), one obtains

$$\int_0^{T'} \varphi_k(t)\, y(t)\, \mathrm{d}t = T_1 \int_0^{T'} \dot{\varphi}_k(t)\, y(t)\, \mathrm{d}t + V \int_0^{T'} \varphi_k(t)\, u(t)\, \mathrm{d}t \tag{3.18}$$

equivalent to (3.16), suitable for identification. Using $N > n_p$ different modulating functions, one obtains an overdetermined set of algebraic equations, which can be solved by, e.g., least-squares. Alternatively, one can use a single MF by considering a receding-horizon version of (3.18), as shown in, e.g., [78]. In addition to the prime reason for avoiding the explicit calculation of time derivatives of measurement signals, the integration has a smoothing effect in measurement noise. As stated in [79], only a few MF have been introduced over time. For example, [77] uses trigonometric functions

$$\varphi_k(t) = \left(\sin\left(\frac{k\pi t}{T'}\right)\right)^k \tag{3.19}$$

while [80] introduces polynomial functions

$$\varphi_k(t) = (T' - t)^k t^k . \tag{3.20}$$

In the literature, further MFs are proposed, whereby [3] gives a good overview. Fourier modulating functions are introduced in [81], while [82] investigates Hermite modulating functions. In [83], trigonometric modulating functions are used, and [84, 85] describes modulating functions based on the Hartley transformation. In [79], one estimates model parameters and system states in finite time using modulating functions. Moreover, the MFM was also extended to fractional-order models by [86]. Other works, e.g. [87], try to build optimal modulation functions adapted to the measurement data to improve the subsequent parameter estimation.

#### 3.1.2.1.3 Poisson Moment Functional (PMF) Approach

In parallel to the development of the MFM, or instead based on its basic idea, different approaches have been developed that avoid the explicit calculation of time derivatives using partial integration [36,79]. One of them is called PMF method, where the explicit computation of the signal derivatives is avoided by partial integration or signal filtering with an exponential kernel function, e.g., [88, 89]. The integral transform

$$M_k\{y(t)\} = \int_0^t \frac{(t-\tau)^k}{k!} e^{-\lambda(t-\tau)} y(\tau)\,\mathrm{d}\tau \tag{3.21}$$

defines the $k$-th order Poisson moment of the signal $y(t)$. Assuming that the support of $y(t)$ is restricted to $[0, \infty[$, this can be interpreted as a modulating function approach using convolution, i.e.,

$$(g_{F_k} * y)(t) = \int_{-\infty}^{\infty} g_{F_k}(\tau)\, y(t-\tau)\,\mathrm{d}t \tag{3.22}$$

with the $k$-th Poisson impulse response $g_{F_k}(t)$ of a stable linear filter with transfer function[1]

$$F_k(s) = \frac{1}{(s+\lambda)^{k+1}}, \quad \lambda > 0 \tag{3.23}$$

[1] Additionally, the filter can be normalized to unity gain by $\lambda^{k+1}$.

and

$$g_{F_k}(t) = \mathcal{L}^{-1}\left\{F_k(s)\right\} = \frac{t^k}{k!}e^{-\lambda t} \, . \tag{3.24}$$

To clarify the notation: the $k-1$-th Poisson impulse function reads

$$g_{F_{k-1}}(t) = \mathcal{L}^{-1}\left\{F_{k-1}(s)\right\} = \mathcal{L}^{-1}\left\{\frac{1}{(s+\lambda)^k}\right\} \, . \tag{3.25}$$

Continuous-time filters are described by transfer functions in the Laplace domain using the complex variable $s$.

For the $i$-th derivative applies

$$M_k\left\{\frac{\mathrm{d}^i}{\mathrm{d}t^i}y(t)\right\} = M_k\left\{y^{(i)}(t)\right\} = \int_0^t \frac{(t-\tau)^k}{k!}e^{-\lambda(t-\tau)}y^{(i)}(\tau)\,\mathrm{d}\tau \, . \tag{3.26}$$

Integration by parts yields

$$\begin{aligned} M_k\left\{y^{(i)}(t)\right\} = &\frac{(t-\tau)^k}{k!}e^{-\lambda(t-\tau)}y^{(i-1)}(\tau)\Big|_0^t - \\ &\int_0^t \left(-\frac{(t-\tau)^{(k-1)}}{(k-1)!}e^{-\lambda(t-\tau)} + \lambda\frac{(t-\tau)^k}{k!}e^{-\lambda(t-\tau)}\right)y^{(i-1)}(\tau)\,\mathrm{d}\tau \end{aligned} \tag{3.27}$$

or

$$M_k\left\{y^{(i)}(t)\right\} = M_{k-1}\left\{y^{(i-1)}(t)\right\} - \lambda M_k\left\{y^{(i-1)}(t)\right\} - g_{F_k}(t)\,y^{(i-1)}(0) \, . \tag{3.28}$$

This property is used repetitively to eliminate all derivatives to obtain a purely algebraic system of equations for parameter identification.

Exemplarily, for the first derivative, one obtains

$$M_k\left\{y^{(1)}(t)\right\} = M_k\left\{\frac{\mathrm{d}}{\mathrm{d}t}y(t)\right\} = M_{k-1}\left\{y(t)\right\} - \lambda M_k\left\{y(t)\right\} - g_{F_k}(t)\,y(0) \; . \tag{3.29}$$

Iteratively, the second derivative reads

$$\begin{aligned} M_k\left\{y^{(2)}(t)\right\} &= -g_{F_k}(t)\,y^{(1)}(0) + M_{k-1}\left\{y^{(1)}(t)\right\} - \lambda M_k\left\{y^{(1)}(t)\right\} \\ &= \lambda^2 M_k\left\{y(t)\right\} - 2\lambda M_{k-1}\left\{y(t)\right\} + M_{k-2}\left\{y(t)\right\} + \\ &\quad \lambda g_{F_k}(t)\,y(0) - g_{F_k}(t)\,y^{(1)}(0) - g_{F_{k-1}}(t)\,y(0) \; . \end{aligned} \tag{3.30}$$

Contrary to the MF approach, the signal's initial values and its time derivatives do not vanish. Hence, the initial values are assumed to be known, are estimated [90], or one waits until the impact of the initial values vanishes due to exponential decay.

To sum up, despite these differences, the PMF method and the MF method are based on the same principles, namely multiplication of the signals with a known differentiable function or kernel and integration.

Again, using the first-order system (3.16) for demonstration purposes, one obtains

$$M_k\left\{y\left(t\right)\right\} = T_1\left(M_{k-1}\left\{y\left(t\right)\right\} - \lambda M_k\left\{y\left(t\right)\right\} - g_{F_k}\left(t\right)y\left(0\right)\right) + VM_k\left\{u\left(t\right)\right\} \tag{3.31}$$

for identification, evaluated at $N > n_p$ discrete points in time, and least-squares solve the resulting algebraic system of equations.

Instead of using (3.28) with filter $F_k\left(s\right)$ iteratively, i.e., Poisson filter chain, one can use filters with different numerator terms. Neglecting initial conditions, applying Laplace transform, and multiplication with $F_k\left(s\right)$, (3.16) yields

$$\begin{aligned} \hat{y}\left(s\right)\frac{1}{\left(s+\lambda\right)^{k+1}} &= -T_1 s\hat{y}\left(s\right)\frac{1}{\left(s+\lambda\right)^{k+1}} + V\hat{u}\left(s\right)\frac{1}{\left(s+\lambda\right)^{k+1}} \\ &= -T_1\hat{y}\left(s\right)\frac{s}{\left(s+\lambda\right)^{k+1}} + V\hat{u}\left(s\right)\frac{1}{\left(s+\lambda\right)^{k+1}}\,. \end{aligned} \tag{3.32}$$

Inverse transform results in

$$\left(g^0_{F_k} * y\right)\left(t\right) = -T_1\left(g^1_{F_k} * y\right)\left(t\right) + V\left(g^0_{F_k} * u\right)\left(t\right) \tag{3.33}$$

with corresponding filters and impulse responses

$$\begin{aligned} F^0_k\left(s\right) &= \frac{1}{\left(s+\lambda\right)^{k+1}}\,, \quad g^0_{F_k}\left(t\right) = \mathcal{L}^{-1}\left\{F^0_k\left(s\right)\right\} \\ F^1_k\left(s\right) &= \frac{s}{\left(s+\lambda\right)^{k+1}}\,, \quad g^1_{F_k}\left(t\right) = \mathcal{L}^{-1}\left\{F^1_k\left(s\right)\right\}\,. \end{aligned} \tag{3.34}$$

Extending (3.22), one obtains

$$\left(g^i_{F_k} * y\right)\left(t\right) = \int_{-\infty}^{\infty} g^i_{F_k}\left(\tau\right)y\left(t-\tau\right)\mathrm{d}t \tag{3.35}$$

where

$$F^i_k\left(s\right) = \frac{s^i}{\left(s+\lambda\right)^{k+1}} \quad \text{and} \quad g^i_{F_k}\left(t\right) = \mathcal{L}^{-1}\left\{F^i_k\left(s\right)\right\}\,, \quad k+1 \geq i\,. \tag{3.36}$$

A significant advantage of the PMF method is that it is particularly suitable for online parameter estimation due to filters to avoid the direct calculation of time derivatives. For this reason, the focus of this thesis is on parameter estimation using PMF, c.f. Section 3.3.

#### 3.1.2.1.4 Integral Equation Approach

The particular case of the above discussed PMF approach with $\lambda = 0$ is called the integral equation approach. Performing repeated integration removes all derivative terms in the differential equation, resulting in an equation suitable for identification. This approach was introduced for linear, nonlinear, and time-varying systems by [91–93]. Studies on the topic of bias and its compensation can be found in, e.g., [94–96].

#### 3.1.2.2 Least-Squares Method and Modifications

The continuous-time (CT) domain methods are the same as those of discrete-time (DT) models, presented in Section 3.1.1.2, Section 3.1.1.3, and Section 3.1.1.4. For this reason, this section is kept short and simple. Only the main differences are pointed out.

In contrast to the DT domain, the least-squares estimation for CT models is asymptotically biased even if the equation error is assumed to be white noise and Gaussian distributed. The main reason for this is the treatment of the time derivatives that changes the noise spectrum, e.g., equation error is no longer uncorrelated [11]. Hence, bias-compensated ordinary least-squares (BC-OLS) approaches, instrumental variables (IV), or statistic approaches are mandatory in case of a significant noise-signal ratio. BC-OLS method for linear CT systems with output noise, and digital filters for eliminating time derivatives, is presented in, e.g., [7]. In [97,98], the system is extended by introducing a known input pre-filter to get knowledge about the estimation bias caused by the integral equation approach or PMF due to colored noise. This approach manages without estimating the noise characteristics. For example, in [99], this method is applied to reduce the estimation bias caused by low-resolution encoders. An optimal IV approach for CT transfer function models is presented in [100].

### 3.1.3 Parameter Identifiability

The term identifiability describes whether the real system can be described with a model identified by a specific identification method. As stated in [2], this property depends on the factors: system, experimental setup, model structure, and identification method. In other words, identifiability is a joint property of an experiment and a model and means that the model parameters are obtained adequately from the experimental data sets [43].

#### 3.1.3.1 Structural Identifiability

As stated in [11], the concept of identifiability includes two different aspects, namely, whether the data sets used for identification contain enough information to distinguish between different model structures and the properties of the model structure itself, i.e., there exists a unique set of parameters or whether different parameter sets lead to equal models. The latter aspect is also known as *structural identifiability* or *deterministic identifiability* [43]. The issue of *structural identifiability* is treated in, e.g., [101], which derives a probabilistic semi-numerical method for testing local structural identifiability for large dynamic systems. An extended observability matrix for local identifiability analysis based on Lie-derivatives is introduced in [102–104] for linear and nonlinear dynamic systems. A sufficient and necessary condition for weakly local observability based on distribution by claiming piecewise-constant input, suited for testing identifiability, is presented in [105, 106] and applied to electrical circuits for arc fault detection in PV-systems [107]. In [2], the parameter identifiability is analyzed and tested for the least-squares method, with the result that structural identifiability leads to requirements for both system and input signal. In [108], one can look up a comprehensive review of identifiability for nonlinear systems.

#### 3.1.3.2 Persistent Excitation

Assuming structural or deterministic identifiability, one must not prevent finding the model parameters by inadequate input signals. To not lose the property of identifiability, suitable input signals must excite all the plant dynamics [43].

Literature, e.g., [11, 43], states that for the estimation of linear transfer functions of order $n$ with $2n$ parameters, the input signal must contain at least $n$ different frequency components. Therefore, a signal suitable for estimating $2n$ parameters is also called persistently excited with order $n$. Exemplarily, one can use n sinusoidal curves. As stated in [10], a signal is persistently exciting of order $n$ if its covariance matrix of order $n$ is positive definite. For example, a pseudorandom binary sequence (PRBS) of length $n$ is persistently exciting or order $n$, while a step function is persistently exciting of order one only.

The asymptotic properties of parameter estimation in linear systems depend only on the input signal's frequency spectrum and not on its time course, e.g., [11]. It follows directly that white noise is an ideal excitation signal for linear systems. PRBS approximates the spectral properties of white noise and is, therefore, a suitable excitation for linear systems, c.f. Section 2.2 or Section 2.2.1.

For nonlinear models, it is advantageous to use input signals containing different amplitudes to excite the system's nonlinearities. In [3], for example, pseudorandom multilevel sequence (PRMS) signals are suggested. A PRMS signal is a time-discrete periodic signal whose time curve shows different amplitudes and whose spectrum is similar to white noise.[2]

## 3.2 Statistical Properties of Parameter Estimation

The quality of a parameter estimate in a statistical sense is mainly determined by the accuracy and quality of the estimated parameter vector. As written down in [3], for example, the accuracy and quality can be characterized by the following properties:

(i) unbiased

(ii) consistent

(iii) efficient

(iv) sufficient

The following definitions are derived from these properties.

**Bias**

**Definition 3.2.** Let $\hat{\mathbf{p}}$ be the estimated value of a nominal parameter vector $\mathbf{p}$ based on an arbitrary number of samples $N$. If the estimation shows a systematic error

$$\mathrm{E}[\hat{\mathbf{p}} - \mathbf{p}] = \mathrm{E}[\hat{\mathbf{p}}] - \mathbf{p} = \Delta\mathbf{p} \neq \mathbf{0} \tag{3.37}$$

this error is called *bias*. Therefore, an unbiased estimation fulfills

$$\mathrm{E}[\hat{\mathbf{p}}] = \mathbf{p} \,. \tag{3.38}$$

[2] PRBS is thus a particular case of the more general term PRMS with two different signal amplitudes, namely $\pm 1$.

**Asymptotic Bias**

**Definition 3.3.** An estimation is called *asymptotically unbiased* if the higher the number of samples $N$, the more accurate the estimation result, i.e., it applies

$$\lim_{N\to\infty} \mathrm{E}[\hat{\mathbf{p}}] = \mathbf{p}\,. \tag{3.39}$$

**Consistency**

**Definition 3.4.** An estimation result $\hat{\mathbf{p}}$ is called *consistent* if it converges to its nominal value $\mathbf{p}$ in probability for $N \to \infty$. According to the definition, the estimate $\hat{\mathbf{p}}$ is consistent if the probability of $P\{\|\hat{\mathbf{p}}-\mathbf{p}\| > \epsilon\}$ meets the condition

$$\lim_{N\to\infty} P\{\|\hat{\mathbf{p}}-\mathbf{p}\| \geq \epsilon\} = 0\,, \quad \epsilon > 0 \tag{3.40}$$

for all admissable $\mathbf{p}$ and all real $\epsilon > 0$. The notation

$$\operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} = \mathbf{p} \tag{3.41}$$

is often used for the different varieties of stochastic convergence, in this particular case for convergence in probability.

**Consistency vs. Asymptotic Unbiasedness**

**Remark 3.1.** Consistency implies asymptotic unbiasedness, but the reverse is not always true. However, consistency and asymptotic unbiasedness are equivalent if the estimator is consistent and asymptotically normally distributed [109]. Assuming this mild regularity conditions *asymptotic unbiasedness* and the so called *consistency* are identical [108, 109]. In this thesis, we will not distinguish between the two terms.

**Efficiency**

**Definition 3.5.** Let $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$ be two unbiased estimates of the parameter vectors $\mathbf{p}$. The efficiency of the estimate $\hat{\mathbf{p}}_1$ compared to the estimate $\hat{\mathbf{p}}_2$ is described by the ratio of its variances $\mathrm{var}(\hat{\mathbf{p}}_1)$ and $\mathrm{var}(\hat{\mathbf{p}}_2)$, respectively.

$$\begin{aligned}\mathrm{var}(\hat{\mathbf{p}}_1) &= \mathrm{E}[(\hat{\mathbf{p}}_1 - \mathbf{p})(\hat{\mathbf{p}}_1 - \mathbf{p})^\top] \\ \mathrm{var}(\hat{\mathbf{p}}_2) &= \mathrm{E}[(\hat{\mathbf{p}}_2 - \mathbf{p})(\hat{\mathbf{p}}_2 - \mathbf{p})^\top]\end{aligned} \tag{3.42}$$

An estimated parameter vector $\hat{\mathbf{p}}_1$ is efficient if there exists no other arbitrary estimation $\hat{\mathbf{p}}_2$, which has a smaller variance, i.e., it applies

$$\forall \hat{\mathbf{p}}_2 \left(\mathrm{var}(\hat{\mathbf{p}}_1) < \mathrm{var}(\hat{\mathbf{p}}_2)\right) . \tag{3.43}$$

**Sufficiency**

**Definition 3.6.** The estimated vector $\hat{\mathbf{p}}$ is sufficient if there exists no other estimate from the same input and output data (measured data), that provide additional information about the nominal parameter vector $\mathbf{p}$. Although sufficient estimates are highly desirable, they can only be realized in particular cases.

The terms presented herein, brevity, characterize the quality of parameter estimation. For a more detailed description of the terms, reference is made to the numerous available literature, for example [110–112].

## 3.3 Online Parameter Estimation Approach Based on Poisson Moment Functional Method and Recursive Least-Squares

For parameter estimation with Poisson moment functionals (PMF) and ordinary least-squares (OLS), we consider strictly proper observable and controllable single-input single-output (SISO) linear time-invariant (LTI) systems of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}(\mathbf{p})\,\mathbf{x} + \mathbf{b}(\mathbf{p})\,u \\ y &= \mathbf{c}^\top(\mathbf{p})\,\mathbf{x}\end{aligned} \tag{3.44}$$

with state vector $\mathbf{x}(t) \in \mathbb{R}^n$, input $u(t) \in \mathbb{R}$, output $y(t) \in \mathbb{R}$, and constant (unknown) parameter vector $\mathbf{p} \in \mathbb{R}^{n_p}$. Moreover, as shown in [113], (3.44) may be represented equivalently in nonlinear observer canonical form

$$y^{(n)}(t) + \sum_{i=0}^{n-1} a_i y^{(i)}(t) = \sum_{j=0}^{m_d} b_j u^{(j)}(t) \tag{3.45}$$

with coefficients $\{a_i, b_j\}$, system order $n > m_d$, and parameter vector

$$\mathbf{p} = \left(a_0, \cdots, a_{n-1}, b_0, \cdots, b_{m_d}\right)^{\mathrm{T}} . \tag{3.46}$$

As described in Section 3.1, different measures for the error between model and process are introduced. Dealing with differential or difference equations, (generalized) equation error methods (EEMs) are usually the methods of choice for parameter estimation to obtain estimation problems linear in the model parameters. Thus, one of the main problems for the identification of CT systems is the elimination of the time derivatives in the occurring signals.[3] One possibility is to transform the differential equations into algebraic equations by multiplicating so-called modulating functions and applying partial integration.

## 3.3.1 Poisson Moment Functional Approach

Let us begin with the most important repeated in brief. As described before, the PMF method is particularly suitable for online parameter estimation. Hence, the integral transform

$$M_k\{y(t)\} = \int_0^t \frac{(t-\tau)^k}{k!} e^{-\lambda(t-\tau)} y(\tau)\,\mathrm{d}\tau \tag{3.47}$$

defines the $k$-th order Poisson moment of the signal $y(t)$ $(k = 0, 1, 2, \ldots)$. Assuming that the support of $y(t)$ is restricted to $[0, \infty[$, this can be interpreted as a modulating function approach using convolution

$$\left(g_{F_k} * y\right)(t) = \int_{-\infty}^{\infty} g_{F_k}(\tau)\, y(t-\tau)\,\mathrm{d}t \tag{3.48}$$

with the impulse response $g_{F_k}(t)$ of a stable linear filter with transfer function[4]

$$F_k(s) = \frac{1}{(s+\lambda)^{k+1}}, \quad \lambda > 0 \tag{3.49}$$

[3] When using so-called OEMs, the problem of eliminating or computing the time-derivatives of signals, is dropped.

[4] Additionally, the filter can be normalized to unity gain by $\lambda^{k+1}$.

and

$$g_{F_k}(t) = \mathcal{L}^{-1}\{F_k(s)\} \ . \tag{3.50}$$

Of course, continuous-time filters are described by transfer functions in the Laplace domain using the complex variable $s$. Integration by parts or Laplace transform shows the effect on the first derivatives[5]

$$M_k\{\dot{y}(t)\} = M_k\left\{\frac{\mathrm{d}}{\mathrm{d}t}y(t)\right\} = M_{k-1}\{y(t)\} - \lambda M_k\{y(t)\} - g_{F_k}(t)\,y(0) \tag{3.51}$$

where initially

$$M_0\{\dot{y}(t)\} = M_0\left\{\frac{\mathrm{d}}{\mathrm{d}t}y(t)\right\} = y(t) - \lambda M_0\{y(t)\} - g_{F_k}(t)\,y(0) \ . \tag{3.52}$$

This property is used repetitively to eliminate all derivatives in (3.45) to obtain a purely algebraic system of equations for parameter identification. Alternatively, one can use (3.35) and (3.36) to avoid filter chains. The initial conditions are assumed to be zero or at least known or are neglected since their impact fades with time depending on the filter time constants. Whenever necessary, the initial conditions may also be estimated as additional parameters.

For each derivative of $y(t)$, one obtains

$$\tilde{y}^{(i)}(t) = \frac{\mathrm{d}^i}{\mathrm{d}t^i}\{\tilde{y}(t)\} = \left(g_F^i * y\right)(t) \tag{3.53}$$

where $(g_F^i * y)(t)$ denotes the convolution of the signal $y(t)$ with the impulse response $g_F^i$ of a stable linear filter $F^i(s)$[6], where

$$F^i(s) = \frac{s^i}{(s+\lambda)^{k+1}} \quad \text{and} \quad g_F^i(t) = \mathcal{L}^{-1}\left\{F^i(s)\right\} \ , \quad k+1 \geq i \, . \tag{3.54}$$

The input signal $u(t)$ and its time derivatives in (3.45) are transformed analogously, resulting in an algebraic system equation for identification

$$\tilde{y}^{(n)}(t) + \sum_{i=0}^{n-1} a_i \tilde{y}^{(i)}(t) = \sum_{j=0}^{m_d} b_j \tilde{u}^{(j)}(t) \tag{3.55}$$

where the accent *tilde* indicates filtered signals.

[5]For a definition of Poisson moment functionals for distributions see [5].

[6]In (3.49), we use the subscript to highlight the order of the PMF moment. This index is replaced to indicate other essential properties, e.g., the filters' discretization type.

### 3.3.2 Implementation Aspects

For implementation on sampled measurement data $t_k = kT$, $1 \leq k \leq N$ with sample time $T$, the filters (3.54) have to be discretized or approximated. Please note that to improve the readability of mathematical expressions, the sampling of signals at discrete time instances $t_k$, e.g., $y(t_k)$, is abbreviated by a subscript $k$, so $y_k$.

One may exactly discretize the plant input filters under the assumption that a zero-order hold generates the inputs. Any other filters operating on continuous-time signals are discretized using bilinear approximation or Tustin's method (trapezoidal rule). More accurate approximations are possible, of course.

Discretization of the filters using zero-order-hold (ZOH) is done with

$$F^i_{zoh}(z) = \frac{z-1}{z}\mathcal{Z}\left\{\left(\mathcal{L}^{-1}\left\{F^i(s)\frac{1}{s}\right\}\right)\Big|_{t=kT}\right\} \tag{3.56}$$

where $z$ indicates the z-transform variable. The corresponding impulse response sequence read

$$g^i_{F,zoh}(k) = \mathcal{Z}^{-1}\left\{F^i_{zoh}(z)\right\} \tag{3.57}$$

allowing the remark that $g^i_{F,zoh}(k) = g^i_{F,zoh;k}$. Bilinear approximation leads to

$$F^i_{tust}(z) = F^i(s') \quad \text{where} \quad s' = \frac{2}{T}\frac{z-1}{z+1} \tag{3.58}$$

with the appropriate impulse response sequence

$$g^i_{F,tust}(k) = \mathcal{Z}^{-1}\left\{F^i_{tust}(z)\right\} . \tag{3.59}$$

Analogously to (3.53), the discrete-time domain applies

$$\tilde{u}^{(j)}_k = \left\{\left(\frac{\mathrm{d}^j}{\mathrm{d}t^j}\{\tilde{u}(t)\}\right)\Big|_{t=kT}\right\} = \left(g^j_{F,zoh} * u\right)(k) \tag{3.60}$$

and

$$\tilde{y}^{(i)}_k = \left\{\left(\frac{\mathrm{d}^i}{\mathrm{d}t^i}\{\tilde{y}(t)\}\right)\Big|_{t=kT}\right\} \approx \left(g^i_{F,tust} * y\right)(k) =: \hat{\tilde{y}}^{(i)}_k \tag{3.61}$$

with the time-discrete convolution

$$\begin{aligned}\left(g^j_{F,zoh} * u\right)(k) &= \sum_{m=-\infty}^{\infty} g^j_{F,zoh;m} u_{k-m} \\ \left(g^i_{F,tust} * y\right)(k) &= \sum_{m=-\infty}^{\infty} g^i_{F,tust;m} y_{k-m} .\end{aligned} \tag{3.62}$$

At this point, it should be noted that the accent *hat* in $\hat{\tilde{y}}_k^{(i)}$ indicates the approximative character. Of course, the quality of the approximation can be improved by increasing the sampling frequency. Indeed, if the signal shape of the continuous-time signals is known, these filters can also be discretized exactly. Therefore, the equation for identification (3.55) applied to sampled measurement data read as

$$\hat{\tilde{y}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{y}}_k^{(i)} = \sum_{j=0}^{m_d} b_j \tilde{u}_k^{(j)} \tag{3.63}$$

with

$$\begin{aligned} \hat{\tilde{y}}_k^{(i)} &= \left(g_{F,tust}^i * y\right)(k)\,, \quad i = 0, 1, \ldots, n \\ \tilde{u}_k^{(j)} &= \left(g_{F,zoh}^j * u\right)(k)\,, \quad j = 0, 1, \ldots, m_d\,. \end{aligned} \tag{3.64}$$

### 3.3.3 Least-Squares Parameter Estimation for Dynamic CT Processes

Resorting (3.63) results in

$$\hat{\tilde{y}}_k^{(n)} = -\sum_{i=0}^{n-1} a_i \hat{\tilde{y}}_k^{(i)} + \sum_{j=0}^{m_d} b_j \tilde{u}_k^{(j)} \tag{3.65}$$

or vector form

$$y_{LSQ;k} = \mathbf{w}_k^\top \mathbf{p} \tag{3.66}$$

with

$$\begin{aligned} y_{LSQ;k} &= \hat{\tilde{y}}_k^{(n)} \\ \mathbf{w}_k^\top &= \left(-\hat{\tilde{y}}_k^{(0)}, \cdots, -\hat{\tilde{y}}_k^{(n-1)}, \tilde{u}_k^{(0)}, \cdots, \tilde{u}_k^{(m_d)}\right) \\ \mathbf{p} &= (a_0, \cdots, a_{n-1}, b_0, \cdots, b_{m_d})^\top \end{aligned} \tag{3.67}$$

where $\mathbf{w}_k^\top \in \mathbb{R}^{n_p}$ denotes the regressor vector, $y_{LSQ;k} \in \mathbb{R}$ is the ordinary least-squares (OLS) output, and $\mathbf{p} \in \mathbb{R}^{n_p}$ is the parameter vector. For $N > n_p$ measurements, (3.65) or (3.66) results in an overdetermined system of equations

$$\underbrace{\begin{pmatrix} y_{LSQ;1} \\ y_{LSQ;2} \\ \vdots \\ y_{LSQ;k} \\ \vdots \\ y_{LSQ;N} \end{pmatrix}}_{=\mathbf{y}_{LSQ}} = \underbrace{\begin{pmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_k^\top \\ \vdots \\ \mathbf{w}_N^\top \end{pmatrix}}_{=\mathbf{W}} \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n_p} \end{pmatrix}}_{=\mathbf{p}} \tag{3.68}$$

where $\mathbf{W} \in \mathbb{R}^{N \times n_p}$, $\mathbf{p} \in \mathbb{R}^{n_p}$ and $\mathbf{y}_{LSQ} \in \mathbb{R}^N$.

For estimating the unknown parameter vector, the OLS algorithm is used. Extending (3.68) by a generalized equation error $v$ results in an underdetermined system of equations

$$\mathbf{y}_{LSQ} = \mathbf{W}\mathbf{p} + \mathbf{v} \tag{3.69}$$

for $\mathbf{p}$ and $\mathbf{v} \in \mathbb{R}^N$. The basic idea of OLS is to find a solution $\hat{\mathbf{p}}$ in such a way that the square norm of the error becomes minimal. Instead of solving the system of linear equations (3.68) directly, it is transformed into an equivalent optimization problem

$$\begin{aligned} \hat{\mathbf{p}} &= \arg \min_{\mathbf{p}} \, ||\mathbf{v}||_2^2 \\ &= \arg \min_{\mathbf{p}} \, ||\mathbf{y}_{LSQ} - \mathbf{W}\mathbf{p}||_2^2 \, . \end{aligned} \tag{3.70}$$

Calculating the gradient concerning the parameters and equate it to zero

$$\begin{aligned} \frac{\partial}{\partial \mathbf{p}} \mathbf{v}^\top \mathbf{v} &= \frac{\partial}{\partial \mathbf{p}} \left(\mathbf{y}_{LSQ} - \mathbf{W}\mathbf{p}\right)^\top \left(\mathbf{y}_{LSQ} - \mathbf{W}\mathbf{p}\right) \\ &= -2\mathbf{y}_{LSQ}\mathbf{W} + 2\mathbf{p}^\top \mathbf{W}^\top \mathbf{W} = \mathbf{0}^\top \end{aligned} \tag{3.71}$$

results directly in

$$\mathbf{W}^\top \mathbf{W}\mathbf{p} = \mathbf{W}^\top \mathbf{y}_{LSQ} \, . \tag{3.72}$$

The invertibility of the matrix $\left(\mathbf{W}^\top \mathbf{W}\right)$ depends on the input and output signals and is often called *persistent excitation*. Assuming regularity of $\left(\mathbf{W}^\top \mathbf{W}\right)$ means $\operatorname{rank}\left(\mathbf{W}^\top \mathbf{W}\right) = n_p$, the optimal solution $\hat{\mathbf{p}}$ in the sense of (3.70) for the unknown parameter vector is

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \tag{3.73}$$

where $\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top$ is called *Moore-Penrose pseudoinverse* of $\mathbf{W}$.[7]

#### 3.3.3.1 Statistical Properties of Ordinary Least-Squares Estimators

Different assumptions are necessary to prove the statistical properties of OLS estimators. The so-called *Gauss-Markov* assumptions (i)-(iv) play an essential role. The *Gauss-Markov* assumptions can be divided into assumptions concerning the specification of the model (i), assumptions concerning the data matrix (ii), (iii), and assumptions concerning the error terms (iv). If all Gauss-Markov assumptions are fulfilled, the OLS estimator is the *best linear unbiased estimator (BLUE)* estimator. A short and elegant proof can be found in [114]. In other words, the estimator is efficient, c.f. Definition 3.5.

(i) **Linearity** The observations $y_{LSQ;k}$ are an affine function of the explanatory variables (regressor or data vector) $\mathbf{w}_k^\top$ and equation error $v_k$, i.e.,

$$y_{LSQ;k} = \mathbf{w}_k^\top \mathbf{p} + v_k \;, \quad k = 1, \ldots, N \;. \tag{3.74}$$

It is essential to note that the linearity assumption refers to the model parameters but not the variables. In addition, it is assumed that the model is correctly specified, i.e., that no relevant explanatory variables are missing or no irrelevant explanatory variables occur.

(ii) **Full column rank** The matrix $\mathbf{W} \in \mathbb{R}^{N \times n_p}$ has full column rank, i.e.,

$$\operatorname{rank}[\mathbf{W}] = n_p \;. \tag{3.75}$$

Firstly, full rank means that the number of observations has to be higher (or at least equal) to the number of parameters $N \geq n_p$, and secondly, the column vectors of the data matrix $\mathbf{W}$ must not have perfect multicollinearity. The rank condition is necessary for identifying the parameters $\mathbf{p}$ of interest and can be understood more broadly as a kind of identification condition.

(iii) **Exogeneity** The explanatory variables and the error variables are stochastically independent, i.e.,

$$\mathrm{E}[v_k|\mathbf{W}] = 0 \;. \tag{3.76}$$

[7] The second-order condition for a minimum requires that the matrix $\mathbf{W}^\top\mathbf{W}$ is positive definite. This condition is valid under very general conditions, provided that $\mathbf{W}$ has full column rank.

This also implies the less strict assumption $\mathrm{E}[v_k] = 0$. In this case, the regressor variables are called *exogenous variables*. For time series, the equation error $v_k$ at a particular time instance $t_k$ must be (stochastically) independent from the values of $\mathbf{W}$ for any time instances. In contrast, regressors that violate this assumption are called *endogenous regressors*.

(iv) **Error terms** The error terms are independent and identically distributed (i.i.d.) with an expected value equal to zero and finite constant variance $\sigma^2$, i.e.,

$$v_k|\mathbf{W} \sim i.i.d.\left(0, \sigma^2\right) . \tag{3.77}$$

This statement can be broken down into three individual assumptions:

- The expected value is equal to zero.

$$\mathrm{E}[v_k] = 0 \tag{3.78}$$

- The error terms have the same finite variance. This property is known as *homoscedasticity*.

$$\mathrm{var}\left(v_k|\mathbf{W}\right) = \sigma^2 , \quad 1 \leq k \leq N \tag{3.79}$$

- The error terms $v_k$ are stochastically independent of each other, i.e.,

$$\mathrm{E}[v_k v_r|\mathbf{W}] = 0 , \quad \text{for } k \neq r . \tag{3.80}$$

(v) **Random sampling** The value pairs $\left\{\mathbf{w}_k^\top, y_{LSQ;k}\right\}$ are obtained by random sampling from a common distribution. If the drawings are independent, the obtained data is independent and identically distributed (i.i.d.).[8]

As mentioned at the beginning of this section, the estimator is called *BLUE* if all *Gauss-Markov* assumptions are fulfilled. However, usually, only a few assumptions are fulfilled, and it must be verified which of the statistical properties mentioned in Section 3.2 are fulfilled by the OLS estimators and which are not.

#### 3.3.3.1.1 Finite Sample Properties

In estimating model parameters, the property of the expected value of the estimation

[8] Please note that this assumption is not one of the Gauss-Markov assumptions. However, since it is necessary assumption to show consistency, it is also mentioned here.

results is of great importance. Inserting (3.69) in (3.73) gives

$$\begin{aligned}\hat{\mathbf{p}} &= \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{y}_{LSQ} = \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\left(\mathbf{W}\mathbf{p}+\mathbf{v}\right)\\ &= \mathbf{p} + \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{v}\end{aligned} \tag{3.81}$$

with the nominal parameter vector $\mathbf{p}$. Taking expectation results in

$$\mathrm{E}[\hat{\mathbf{p}}] = \mathrm{E}[\mathbf{p} + \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{y}_{LSQ}] = \mathbf{p} + \mathrm{E}[\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{v}] = \mathbf{p} + \Delta\mathbf{p} \tag{3.82}$$

where the second summand $\Delta\mathbf{p} = \mathrm{E}[\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{v}]$ terms the deviation between the expected value of the estimated and nominal parameter vector $\Delta\mathbf{p} = \mathrm{E}[\hat{\mathbf{p}}] - \mathbf{p}$, commonly known as *bias*. Let us assume for the moment that we are looking at a static process. For deterministic regressor vectors $\mathbf{w}_k^\top$ or regressor matrices $\mathbf{W}$, and if *Gauss-Markov* assumptions (i)-(iii) are fulfilled, it applies

$$\mathrm{E}[\hat{\mathbf{p}}] = \mathbf{p} + \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\,\mathrm{E}[\mathbf{v}] = \mathbf{p}\,. \tag{3.83}$$

In this case, the estimation result is unbiased, even for a finite number of samples. For stochastic regressors, the concept of the expected value must be extended. Assuming a fixed realization of the random data matrix, one calls the expected value the conditional expected value. Hence, the conditional expectation of the bias-term, assuming a present data matrix $\mathbf{W}$, is given by

$$\mathrm{E}[\Delta\mathbf{p}|\mathbf{W}] = \mathrm{E}[\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\mathbf{v}|\mathbf{W}]\,. \tag{3.84}$$

Let us assume that the data matrix $\mathbf{W}$ and the equation error $\mathbf{v}$ are stochastically independent. This assumption implies that the elements of the data matrix $w_{i,j} \in \mathbf{W}$ are uncorrelated with the error terms $v_i$. Then, the conditional expectation value applies

$$\mathrm{E}[\Delta\mathbf{p}|\mathbf{W}] = \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\,\mathrm{E}[\mathbf{v}]\,. \tag{3.85}$$

With $\mathrm{E}[v_k] = 0$ the conditional expectation value for the bias-term is

$$\mathrm{E}[\Delta\mathbf{p}|\mathbf{W}] = \mathbf{0}\,. \tag{3.86}$$

Thus, the conditional expected value does not depend on the data matrix, which means that the (unconditional) expectation value also becomes zero, i.e.,

$$\mathrm{E}[\Delta\mathbf{p}] = \mathbf{0} \tag{3.87}$$

resulting again in an *unbiased* estimation. Once more, this applies to a finite number of samples. Alternatively, by assuming that the data matrix $\mathbf{W}$ and the equation error $\mathbf{v}$ are stochastically independent, the expected value in (3.82) can be split, and it applies

$$\mathrm{E}[\hat{\mathbf{p}}] = \mathbf{p} + \mathrm{E}[\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top]\,\mathrm{E}[\mathbf{v}] = \mathbf{p} \tag{3.88}$$

assuming $\mathrm{E}[v_k] = 0$.

Apart from the two exceptional cases just mentioned, for dependent regressor and equation error, the expectation value is generally non-zero. Hence, the estimation result becomes biased, at least for a finite number of samples.

#### 3.3.3.1.2 Asymptotic Properties

In control engineering tasks, usually dynamic processes with noise corrupted outputs are treated. Consequently, the data matrix contains, among other things, the noisy output signals and their time-derivatives. Intuitively, the regressor and equation error are dependent, and the estimation bias will no longer be zero for a finite number of samples, e.g., [2, 3, 9, 43]. However, if the *Gauss-Markov* assumptions (i)-(iv) are fulfilled, and additionally the assumption (v) applies, then in favorable cases, at least the asymptotic unbiasedness (consistency) of the estimator can be proved, using the law of large numbers and *convergence in probability (plim)*. A short introduction to stochastic convergence concepts is given in Section A.3.6. Applying *plim* to (3.81), one obtains

$$\operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} = \operatorname*{plim}_{N\to\infty} \left(\mathbf{p} + \left(\mathbf{W}^\top\mathbf{W}\right)\mathbf{W}^\top\mathbf{v}\right) = \mathbf{p} + \operatorname*{plim}_{N\to\infty} \left(\mathbf{W}^\top\mathbf{W}\right)\mathbf{W}^\top\mathbf{v}\,. \tag{3.89}$$

The impact of correlated data and error terms on the consistency of the estimation result is investigated in detail in Section 3.3.4.

### 3.3.4 Bias Problematic with Noisy Sampled System Outputs

The output of a continuous-time system is measured at various discrete instances of time $t_k = kT$, $1 \leq k \leq N$, where $N$ is the total number of samples. The measured samples

$$y_m(t_k) = y_{m;k} = y_k + e_k, \quad 1 \leq k \leq N \tag{3.90}$$

are assumed to be disturbed by an additive random noise sequence $e_k$. If not specified more precisely, relatively broad assumptions apply to the noise term, $e_k$. One assumes that the random variable $e_k$ is a realization of an ergodic, i.e., stationary, white noise process with its expected value $\mathrm{E}[e_k] = 0$ and constant variance $\sigma^2$. To work out the bias problem, one can retain the measured samples or the filtered signals in the equation for identification. For example, inserting (3.90) in (3.63) and using (3.60) or (3.61) results in

$$\hat{\tilde{y}}_{m;k}^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{y}}_{m;k}^{(i)} = \sum_{j=0}^{m_d} b_j \tilde{u}_k^{(j)} + \underbrace{\hat{\tilde{e}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{e}}_k^{(i)}}_{v_k} \tag{3.91}$$

with

$$\begin{aligned} \tilde{u}_k^{(j)} &= \left\{ \left( \frac{\mathrm{d}^j}{\mathrm{d}t^j} \left\{ \tilde{u}(t) \right\} \right) \Big|_{t=kT} \right\} = \left( g_{F,zoh}^j * u \right)(k) \\ \tilde{y}_{m;k}^{(i)} &= \left\{ \left( \frac{\mathrm{d}^i}{\mathrm{d}t^i} \left\{ \tilde{y}_m(t) \right\} \right) \Big|_{t=kT} \right\} \approx \left( g_{F,tust}^i * y_m \right)(k) =: \hat{\tilde{y}}_{m;k}^{(i)} \end{aligned} \tag{3.92}$$

and

$$\tilde{e}_k^{(i)} = \left\{ \left( \frac{\mathrm{d}^i}{\mathrm{d}t^i} \left\{ \tilde{e}(t) \right\} \right) \Big|_{t=kT} \right\} \approx \left( g_{F,tust}^i * e \right)(k) =: \hat{\tilde{e}}_k^{(i)} \tag{3.93}$$

where $v_k$ is a composite noise term, called *filtered white noise*.[9] The term *filtered* in this context refers to the filtering effect by the plant dynamics and not the PMF to eliminate the time derivatives. A similar challenge arises in the parameter estimation of discrete-time transfer functions $G(z)$. Due to the plant dynamics filtering effect, the estimation result is biased, even if $e_k$ is stationary with $\mathrm{E}[e_k] = 0$. A common requirement for a bias-free estimation of DT transfer function parameters is that the noise term must be generated from white noise by a filter with transfer function $\frac{1}{A(z)}$, e.g., [2, 3]. Unfortunately, this does not apply in the case of CT, and the estimation result remains biased. For more detailed information, please refer to Section 3.3.5.1.

[9] Again, plant input filters are discretized by zero order hold in an exact manner, while other filters are discretized by bilinear approximation, see Section 3.3.2

For $N$ samples, one obtains

$$\underbrace{\begin{pmatrix} y_{LSQ;1} \\ y_{LSQ;2} \\ \vdots \\ y_{LSQ;k} \\ \vdots \\ y_{LSQ;N} \end{pmatrix}}_{=\mathbf{y}_{LSQ}} = \underbrace{\begin{pmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_k^\top \\ \vdots \\ \mathbf{w}_N^\top \end{pmatrix}}_{=\mathbf{W}} \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n_p} \end{pmatrix}}_{=\mathbf{p}} + \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \\ \vdots \\ v_N \end{pmatrix}}_{=\mathbf{v}} \tag{3.94}$$

or in more compact form

$$\mathbf{y}_{LSQ} = \mathbf{W}\mathbf{p} + \mathbf{v} \tag{3.95}$$

where $\mathbf{y}_{LSQ} \in \mathbb{R}^N$, $\mathbf{W} \in \mathbb{R}^{N \times n_p}$, $\mathbf{v} \in \mathbb{R}^N$, with the abbreviations for one single measurement

$$\begin{aligned} y_{LSQ;k} &= \hat{\tilde{y}}_{m;k}^{(n)} \\ \mathbf{w}_k^\top &= \begin{pmatrix} -\hat{\tilde{y}}_{m;k}^{(0)} & \cdots & -\hat{\tilde{y}}_{m;k}^{(n-1)} & \tilde{u}_k^{(0)} & \cdots & \tilde{u}_k^{(m_d)} \end{pmatrix} \\ v_k &= \hat{\tilde{e}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{e}}_k^{(i)} \end{aligned} \tag{3.96}$$

and

$$\mathbf{v} = \begin{pmatrix} v_1 & \cdots & v_k & \cdots & v_N \end{pmatrix}^\top, \quad \mathbf{v} \in \mathbb{R}^N . \tag{3.97}$$

Inserting (3.95) in (3.73) to obtain the optimal solution in the least-squares sense ends up with

$$\begin{aligned} \hat{\mathbf{p}} &= \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{W}\mathbf{p} + \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{v} \\ &= \mathbf{p} + \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{v} \end{aligned} \tag{3.98}$$

where $\mathbf{p}$ is the nominal parameter vector, and

$$\mathbf{P} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \tag{3.99}$$

is the parameter covariance matrix.

As described in the previous section, the estimation result's expectation value is an essential property of an estimator. However, we expect that the estimation results are no longer unbiased due to measurement noise and dynamical models. A glance at relevant technical literature, e.g., [2, 3, 43], shows how or why regressor and error terms are correlated with each other in the parameter estimation of continuous-time dynamic systems using OLS. For example, [43] notes that the bias analysis quickly becomes impracticable due to the necessary evaluation of the expected value of relatively complex functions. Nevertheless, in [2], the necessary and sufficient condition

$$\mathrm{E}[\mathbf{W}^\top \mathbf{v}] = 0 \tag{3.100}$$

for unbiased estimation of the unknown parameters of a CT system is introduced. The other way around, if this condition is not fulfilled, the estimation result is biased for a finite number of samples. Unfortunately, no further analysis of the bias, e.g., its size, can be made if $\mathbf{W}^\top \mathbf{v} \neq \mathbf{0}$. One can circumvent the arising challenge by considering the asymptotic bias. For this purpose, we analyze the term $\mathrm{E}[\mathbf{W}^\top \mathbf{v}]$ in more detail. Inserting

$$\mathbf{W}^\top = \begin{pmatrix} \mathbf{W}_y^\top \mathbf{v} \\ \hline \mathbf{W}_u^\top \mathbf{v} \end{pmatrix} = \begin{pmatrix} -\hat{\tilde{y}}_{m;1}^{(0)} & -\hat{\tilde{y}}_{m;2}^{(0)} & \cdots & -\hat{\tilde{y}}_{m;N}^{(0)} \\ -\hat{\tilde{y}}_{m;1}^{(1)} & -\hat{\tilde{y}}_{m;2}^{(1)} & \cdots & -\hat{\tilde{y}}_{m;N}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{\tilde{y}}_{m;1}^{(n-1)} & -\hat{\tilde{y}}_{m;2}^{(n-1)} & \cdots & -\hat{\tilde{y}}_{m;N}^{(n-1)} \\ \hline u_1^{(0)} & u_2^{(0)} & \cdots & u_N^{(0)} \\ u_1^{(1)} & u_2^{(1)} & \cdots & u_N^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{(m_d)} & u_2^{(m_d)} & \cdots & u_N^{(m_d)} \end{pmatrix} \tag{3.101}$$

results in

$$\mathrm{E}[\mathbf{W}^\top \mathbf{v}] = \mathrm{E}\left[\begin{pmatrix} -\hat{\tilde{y}}_{m;1}^{(0)} v_1 - \hat{\tilde{y}}_{m;2}^{(0)} v_2 - \cdots - \hat{\tilde{y}}_{m;N}^{(0)} v_N \\ -\hat{\tilde{y}}_{m;1}^{(1)} v_1 - \hat{\tilde{y}}_{m;2}^{(1)} v_2 - \cdots - \hat{\tilde{y}}_{m;N}^{(1)} v_N \\ \vdots \\ -\hat{\tilde{y}}_{m;1}^{(n-1)} v_1 - \hat{\tilde{y}}_{m;2}^{(n-1)} v_2 - \cdots - \hat{\tilde{y}}_{m;N}^{(n-1)} v_N \\ \hline u_1^{(0)} v_1 + u_2^{(0)} v_2 + \cdots + u_N^{(0)} v_N \\ u_1^{(1)} v_1 + u_2^{(1)} v_2 + \cdots + u_N^{(1)} v_N \\ \vdots \\ u_1^{(m_d)} v_1 + u_2^{(m_d)} v_2 + \cdots + u_N^{(m_d)} v_N \end{pmatrix}\right] = \begin{pmatrix} -\mathrm{E}[\sum_{k=1}^N \hat{\tilde{y}}_{m;k}^{(0)} v_k] \\ -\mathrm{E}[\sum_{k=1}^N \hat{\tilde{y}}_{m;k}^{(1)} v_k] \\ \vdots \\ -\mathrm{E}[\sum_{k=1}^N \hat{\tilde{y}}_{m;k}^{(n-1)} v_k] \\ \hline \mathrm{E}[\sum_{k=1}^N \tilde{u}_k^{(0)} v_k] \\ \mathrm{E}[\sum_{k=1}^N \tilde{u}_k^{(1)} v_k] \\ \vdots \\ \mathrm{E}[\sum_{k=1}^N \tilde{u}_k^{(m_d)} v_k] \end{pmatrix}. \tag{3.102}$$

Taking into account, the response of DT LTI systems with random data (A.81), e.g., [115–118], and the common assumption $\mathrm{E}[e_k] = 0$ one obtains

$$\mathrm{E}[\hat{\tilde{e}}_k^{(i)}] = \mathrm{E}[e_k] \sum_{k=1}^{\infty} g_{F,tust;k}^{i} = 0\,. \tag{3.103}$$

Supplementary, assuming noise-free inputs (open-loop control), i.e., deterministic inputs, (3.102) simplifies to

$$\mathrm{E}[\mathbf{W}^\top \mathbf{v}] = \mathrm{E}[\begin{pmatrix} \mathbf{W}_y^\top \mathbf{v} \\ \mathbf{W}_u^\top \mathbf{v} \end{pmatrix}] = \begin{pmatrix} -\mathrm{E}[\sum_{k=1}^{N} \hat{\tilde{y}}_{m;k}^{(0)} v_k] \\ -\mathrm{E}[\sum_{k=1}^{N} \hat{\tilde{y}}_{m;k}^{(1)} v_k] \\ \vdots \\ -\mathrm{E}[\sum_{k=1}^{N} \hat{\tilde{y}}_{m;k}^{(n-1)} v_k] \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \mathrm{E}[\mathbf{W}_y^\top \mathbf{v}] \\ \mathbf{0} \end{pmatrix}\,. \tag{3.104}$$

Picking out the $j$-th row of $\mathbf{W}_y^\top$, inserting (3.90) and the composite noise term $v_k$ from (3.96) results in

$$\begin{aligned} -\mathrm{E}[\sum_{k=1}^{N} \hat{\tilde{y}}_{m;k}^{(j)} v_k] &= -\mathrm{E}[\sum_{k=1}^{N} \left(\hat{\tilde{y}}_k^{(j)} + \hat{\tilde{e}}_k^{(j)}\right) \left(\hat{\tilde{e}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{e}}_k^{(i)}\right)] \\ &= -\sum_{k=1}^{N} \hat{\tilde{y}}_k^{(j)}\, \mathrm{E}[\hat{\tilde{e}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{e}}_k^{(i)}] - \sum_{k=1}^{N} \left(\mathrm{E}[\hat{\tilde{e}}_k^{(j)} \hat{\tilde{e}}_k^{(n)}] + \sum_{i=0}^{n-1} a_i\, \mathrm{E}[\hat{\tilde{e}}_k^{(j)} \hat{\tilde{e}}_k^{(i)}]\right)\,. \end{aligned} \tag{3.105}$$

Using (3.103) once more, the first term in (3.105) vanishes, and it applies

$$\begin{aligned} -\mathrm{E}[\sum_{k=1}^{N} \hat{\tilde{y}}_{m;k}^{(j)} v_k] &= -\sum_{k=1}^{N} \left(\mathrm{E}[\hat{\tilde{e}}_k^{(j)} \hat{\tilde{e}}_k^{(n)}] + \sum_{i=0}^{n-1} a_i\, \mathrm{E}[\hat{\tilde{e}}_k^{(j)} \hat{\tilde{e}}_k^{(i)}]\right) \\ &= -\sum_{k=1}^{N} \left(r_{\hat{\tilde{e}}^{(j)} \hat{\tilde{e}}^{(n)}}(0) + \sum_{i=0}^{n-1} a_i r_{\hat{\tilde{e}}^{(j)} \hat{\tilde{e}}^{(i)}}(0)\right) \end{aligned} \tag{3.106}$$

with $(i,j) = \{0, 1, \ldots, n{-}1\}$. Derivative (stationary) stochastic processes are treated in [2,119,120]. One can show that for the autocorrelation function (ACF) of a stationary random process $x(t)$ and associative time derivatives $x^{(i)}(t)$ or $x^{(j)}(t)$ applies

$$r_{x^{(j)} x^{(i)}}(\tau) = \mathrm{E}[x^{(j)}(t)\, x^{(i)}(t+\tau)] = (-1)^j r_{xx}^{(j+i)}(\tau) = (-1)^j \frac{\mathrm{d}^{j+i}}{\mathrm{d}\tau^{j+i}} r_{xx}(\tau)\,. \tag{3.107}$$

Using (A.57)

$$r_{x^{(j)}x^{(i)}}(0) = \begin{cases} (-1)^{\frac{j+i}{2}}\, \mathrm{E}[\left(\hat{\hat{e}}_k^{\left(\frac{j+i}{2}\right)}\right)^2], & j+i = \{0,2,4,\dots\} \\ 0\,, & j+i = \{1,3,5,\dots\} \end{cases} \tag{3.108}$$

whereby

$$\mathrm{E}[\left(\hat{\hat{e}}_k^{\left(\frac{j+i}{2}\right)}\right)^2] > 0\,, \quad j+i = \{0,2,4,\dots\} \tag{3.109}$$

can be interpreted as the average signal power of the respective signals, one obtains

$$\mathrm{E}[\mathbf{W}^\top \mathbf{v}] \neq \mathbf{0} \tag{3.110}$$

resulting in a biased estimation result for a small (finite) number of samples.

For more in-depth investigations, e.g., estimating the size of the bias for compensation purposes, asymptotic properties are necessary. As stated in Definition 3.4, an estimator is consistent, means asymptotically unbiased, if the bias of the estimation result $\hat{\mathbf{p}}$ converges in probability towards zero for $N \to \infty$. To check (3.98) for consistency or asymptotic unbiasedness, respectively, one applies

$$\operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} = \operatorname*{plim}_{N\to\infty} \left(\mathbf{p} + \mathbf{P}\mathbf{W}^\top \mathbf{v}\right) \tag{3.111}$$

where plim denotes the convergence in probability as introduced in Section A.3.6.

From Definition A.1, one can derive the calculation rules follow needed to calculate the asymptotic bias, see Corollary A.1. For better readability, one recalls it here:

**Convergence in Probability: Calculation Rules**

**Corollary 3.1.** Let $X_n$, $Y_n$ be univariate random variables with observations $\{X_1, X_2, \dots\}$ and $\{Y_1, Y_2, \dots\}$. Assuming $\operatorname{plim}_{n\to\infty} X_n = a$ and $\operatorname{plim}_{n\to\infty} Y_n = b$, then applies:

$$\begin{aligned} \operatorname*{plim}_{n\to\infty}(X_n \pm Y_n) &= \operatorname*{plim}_{n\to\infty}(X_n) \pm \operatorname*{plim}_{n\to\infty}(Y_n) = a \pm b\,, \\ \operatorname*{plim}_{n\to\infty}(X_n Y_n) &= \operatorname*{plim}_{n\to\infty}(X_n) \operatorname*{plim}_{n\to\infty}(Y_n) = ab\,, \\ \operatorname*{plim}_{n\to\infty}\left(\frac{X_n}{Y_n}\right) &= \frac{\operatorname{plim}_{n\to\infty}(X_n)}{\operatorname{plim}_{n\to\infty}(Y_n)} = \frac{a}{b}\,, \text{ if } b \neq 0\,. \end{aligned} \tag{3.112}$$

If $\mathbf{A}$ is an adequate and nonsingular matrix of random variables, then:

$$\underset{n\to\infty}{\text{plim}}\left(\mathbf{A}_n^{-1}\right) = \left(\underset{n\to\infty}{\text{plim}}\left(\mathbf{A}_n\right)\right)^{-1} \tag{3.113}$$

Applying the calculation rule for convergence in probability (A.64) or (3.112) to (3.111), using (3.99) and taking into account that $\mathbf{p}$ is deterministic, one obtains

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}}\,\hat{\mathbf{p}} &= \mathbf{p} + \left(\underset{N\to\infty}{\text{plim}}\,\mathbf{P}\mathbf{W}^\top\mathbf{v}\right) \\
&= \mathbf{p} + \underset{N\to\infty}{\text{plim}}\left(N\mathbf{P}\frac{1}{N}\mathbf{W}^\top\mathbf{v}\right) \\
&= \mathbf{p} + \underset{N\to\infty}{\text{plim}}\left(\left(\frac{1}{N}\mathbf{W}^\top\mathbf{W}\right)^{-1}\right)\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\mathbf{W}^\top\mathbf{v}\right) .
\end{aligned} \tag{3.114}$$

Using the calculation rule for stochastic matrices and vectors from (A.65) or (3.113), (3.114) can be rewritten to

$$\underset{N\to\infty}{\text{plim}}\,\hat{\mathbf{p}} = \mathbf{p} + \left(\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\mathbf{W}^\top\mathbf{W}\right)\right)^{-1}\underset{N\to\infty}{\text{plim}}\left(\frac{1}{N}\mathbf{W}^\top\mathbf{v}\right) . \tag{3.115}$$

Let us first deal with the matrix $\mathbf{W}^\top\mathbf{W} \in \mathbb{R}^{n_p\times n_p}$. The square matrix $\mathbf{W}^\top\mathbf{W}$ reads in detail

$$\mathbf{W}^\top\mathbf{W} = \begin{pmatrix}
\sum_{i=1}^N w_{i,1}^2 & \sum_{i=1}^N w_{i,1}w_{i,2} & \cdots & \sum_{i=1}^N w_{i,1}w_{i,n_p} \\
\sum_{i=1}^N w_{i,2}w_{i,1} & \sum_{i=1}^N w_{i,1}^2 & \cdots & \sum_{i=1}^N w_{i,2}w_{i,n_p} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^N w_{i,n_p}w_{i,1} & \sum_{i=1}^N w_{i,n_p}w_{i,2} & \cdots & \sum_{i=1}^N w_{i,n_p}^2
\end{pmatrix} \tag{3.116}$$

with

$$\mathbf{W} = \begin{pmatrix}
w_{1,1} & w_{1,2} & \cdots & w_{1,n_p} \\
w_{2,1} & w_{2,2} & \cdots & w_{2,n_p} \\
\vdots & \vdots & \ddots & \vdots \\
w_{N,1} & \cdots & \cdots & w_{N,n_p}
\end{pmatrix} . \tag{3.117}$$

Each element of the square matrix $\mathbf{W}^\top\mathbf{W}$ is a sum of $N$ random variables $\sum_{i=1}^N w_{i,h}w_{i,g}$, $1 \leq (g,h) \leq n_p$, and thus again, a random variable. However, its average value for an infinite number of samples

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} w_{i,h} w_{i,g} = q_{h,g},\ 1 \leq (g,h) \leq n_p \tag{3.118}$$

converges against a fixed number $q_{h,g}$. Therefore it applies

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \mathbf{W}^\top \mathbf{W} = \mathbf{Q}_W \tag{3.119}$$

where $\mathbf{Q}_W \in \mathbb{R}^{n_p \times n_p}$ is a non-stochastic matrix with full rank. Using this property, (3.115) simplifies to

$$\operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} = \mathbf{p} + \mathbf{Q}_W^{-1} \operatorname*{plim}_{N\to\infty} \left( \frac{1}{N} \mathbf{W}^\top \mathbf{v} \right) = \mathbf{p} + \Delta\mathbf{p} \tag{3.120}$$

and it is evident that the estimation result in (3.98) or (3.120) is consistent or asymptotically unbiased only if applies

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \mathbf{W}^\top \mathbf{v} = \mathbf{0}\,. \tag{3.121}$$

To derive a term for the asymptotic bias, the (weak) law of large numbers is necessary.

**Weak Law of Large Numbers (Bernoulli's Theorem)**

**Theorem 3.1.** Briefly, the weak law of large numbers or Bernoulli's theorem states that for a sequence of independent and identically distributed random variables $X_n$, the sample mean $\bar{X}_n$ tends to the population mean $\mathrm{E}[X]$ as the sample size $N$ increases.
Let $X_n = \{X_1, X_2, \dots\}$ be a sequence of i.i.d. random variables with finite sample mean $\bar{X}_n = \mathrm{E}[X_n]$. Then it applies:

$$P\left( |\bar{X}_n - \mathrm{E}[X]| < \epsilon \right) \geq 1 - \delta \tag{3.122}$$

where $\epsilon > 0$ and $0 < \delta < 1$ are small numbers. Consequently it applies:

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i = \mathrm{E}[X]\,. \tag{3.123}$$

However, for time series, often, the requirement of uncorrelated samples is not given. Thus, in [121], a more applicable definition for weak stationary random processes is stated.

**Weak Law of Large Numbers for Weak Stationary Processes**

**Lemma 3.1.** Let $\{X_1, X_2, \ldots\}$ be a wide (weak) sense stationary (WSS) process with absolutely summable auto-covariances. Then:

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i = \mathrm{E}[X] \tag{3.124}$$

Assuming an ergodic, i.e., stationary, stochastic process, and taking into account the fact that a filtered WSS process is also a WSS process, see [115, 116], one can apply Lemma 3.1 to (3.121), resulting in

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \mathbf{W}^\top \mathbf{v} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbf{w}_k v_k = \mathrm{E}[\mathbf{w}_k v_k] = \mathbf{0} \tag{3.125}$$

for asymptotically unbiased or consistent estimates, where $\mathbf{w}_k \in \mathbb{R}^{n_p}$ is the $k$-th column of $\mathbf{W}^\top \in \mathbb{R}^{n_p \times N}$.

**Asymptotic Bias**

**Result 3.1.** In general, the term $\mathrm{E}[\mathbf{w}_k v_k] \neq \mathbf{0}$, whereby the estimation result is asymptotically biased and the asymptotic bias $\Delta\mathbf{p}$ is

$$\Delta\mathbf{p} = \mathbf{Q}_W^{-1} \mathrm{E}[\mathbf{w}_k v_k] = N \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \, \mathrm{E}[\mathbf{w}_k v_k] \, . \tag{3.126}$$

**Asymptotically Unbiased Estimation**

**Result 3.2.** Finally, for the asymptotically unbiased parameter vector one obtains

$$\mathbf{p} = \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - N \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \, \mathrm{E}[\mathbf{w}_k v_k] = \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - \Delta\mathbf{p} \, . \tag{3.127}$$

where $\hat{\mathbf{p}}$ and $\mathbf{P}$ are obtained by ordinary least-squares (OLS) estimation

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \, . \tag{3.128}$$

### 3.3.5 Analytical Solution for the Estimation Bias Caused by PMF Approach

As previously discussed, the CT differential equation in nonlinear observer canonical form is sampled at a discrete time instance

$$y_{m;k}^{(n)} + \sum_{i=0}^{n-1} a_i y_{m;k}^{(i)} = \sum_{j=0}^{m_d} b_j u_k^{(j)} + e_k^{(n)} + \sum_{i=0}^{n-1} a_i e_k^{(i)} \tag{3.129}$$

where $e_k$ denotes the additive random output noise. We use the PMF method to avoid direct computation of time derivatives of measured input- and output signals. By proceeding as described in Section 3.3.1, one obtains

$$\begin{aligned} \tilde{u}_k^{(j)} &= \left\{ \left( \frac{\mathrm{d}^j}{\mathrm{d}t^j} \left\{ \tilde{u}(t) \right\} \right) \bigg|_{t=kT} \right\} &&= \left( g_{F,zoh}^j * u \right)(k) \\ \tilde{y}_{m;k}^{(i)} &= \left\{ \left( \frac{\mathrm{d}^i}{\mathrm{d}t^i} \left\{ \tilde{y}_m(t) \right\} \right) \bigg|_{t=kT} \right\} &&\approx \left( g_{F,tust}^i * y_m \right)(k) =: \hat{\tilde{y}}_{m;k}^{(i)} \end{aligned} \tag{3.130}$$

and

$$\tilde{e}_k^{(i)} = \left\{ \left( \frac{\mathrm{d}^i}{\mathrm{d}t^i} \left\{ \tilde{e}(t) \right\} \right) \bigg|_{t=kT} \right\} \approx \left( g_{F,tust}^i * e \right)(k) =: \hat{\tilde{e}}_k^{(i)} \tag{3.131}$$

with CT filters

$$F^i(s) = \frac{s^i}{(1+sT_f)^n} \quad \text{and} \quad g_F^i(t) = \mathcal{L}^{-1}\left\{ F^i(s) \right\}, \quad 0 \le i \le n \tag{3.132}$$

and appropriately discretized filters $F_{zoh}^i(z), F_{tust}^i(z)$ with the corresponding impulse responses $g_{F,zoh}^i(k), g_{F,tust}^i(k)$. Note that the system order determines the filter order, and $i$ labels the respecting time derivative. Thus, finally, one obtains a purely algebraic equation, suitable for identification, i.e.,

$$\begin{aligned} y_{LSQ;k} &= \hat{\tilde{y}}_{m;k}^{(n)} \\ \mathbf{w}_k^\top &= \begin{pmatrix} -\hat{\tilde{y}}_{m;k}^{(0)} & \cdots & -\hat{\tilde{y}}_{m;k}^{(n-1)} & \tilde{u}_k^{(0)} & \cdots & \tilde{u}_k^{(m_d)} \end{pmatrix} \\ v_k &= \hat{\tilde{e}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\tilde{e}}_k^{(i)} . \end{aligned} \tag{3.133}$$

For implementation purposes, the filters must be discretized. It should be noted that the type of discretization is not discussed here. More detailed considerations on this topic are given in Section 3.3.2 and Section 3.4. One obtains the appropriate discretized filter

$$F^i(z) = \frac{\sum_{j=0}^{n} h_j^i z^j}{1 + \sum_{j=1}^{n} d_j z^j} = \frac{\sum_{j=0}^{n} h_j^i z^j}{d(z)} \tag{3.134}$$

with numerator coefficients $h_j^i$ and common denominator $d(z)$. Equivalent to (3.134), dividing by $z^n$ results in a causal infinite impulse response (IIR) filter

$$F^i\left(z^{-1}\right) = \frac{\sum_{j=0}^{n} h_j^i z^{-j}}{1 + \sum_{j=1}^{n} d_j z^{-j}} = \frac{\sum_{j=0}^{n} h_j^i z^{-j}}{d\left(z^{-1}\right)} \tag{3.135}$$

where $z^{-1}$ is just the inverse of $z$.

In order to calculate the estimation bias $\Delta\mathbf{p}$, a calculation rule for the term $\mathrm{E}[\mathbf{w}_k v_k]$ has to be derived. As already shown in (3.104), the data vector $\mathbf{w}_k$ can be split up into a stochastic part $\mathbf{w}_{y;k}$ containing (random) noisy outputs, and a deterministic part $\mathbf{w}_{u;k}$ including the undisturbed inputs. Assuming an ergodic noise process $e_k$ with $\mathrm{E}[e_k] = 0$, and taking into account the response of LTI systems to random input data, e.g., (A.81), one obtains

$$\mathrm{E}[\hat{\hat{e}}_k^{(i)}] = \mathrm{E}[e_k] \sum_{k=1}^{\infty} g_{F,tust;k}^{i} = 0 \, . \tag{3.136}$$

Considering

$$v_k = \hat{\hat{e}}_k^{(n)} + \sum_{i=0}^{n-1} a_i \hat{\hat{e}}_k^{(i)} \tag{3.137}$$

we finally get

$$\mathrm{E}[\mathbf{w}_{u;k} v_k] = \mathbf{w}_{u;k}\, \mathrm{E}[v_k] = 0 \, . \tag{3.138}$$

Hence, the term $\mathrm{E}[\mathbf{w}_k v_k]$ simplifies to

$$\mathrm{E}[\mathbf{w}_k v_k] = \mathrm{E}[\begin{pmatrix} \mathbf{w}_{y;k} v_k \\ \mathbf{w}_{u;k} v_k \end{pmatrix}] = \mathrm{E}[\begin{pmatrix} \mathbf{w}_{y;k} v_k \\ \mathbf{0} \end{pmatrix}] = \begin{pmatrix} \mathrm{E}[\mathbf{w}_{y;k} v_k] \\ \mathbf{0} \end{pmatrix} \tag{3.139}$$

where

$$\mathbf{w}_{y;k} = \begin{pmatrix} -\hat{\hat{y}}_{m;k}^{(0)} \\ \vdots \\ -\hat{\hat{y}}_{m;k}^{(n-1)} \end{pmatrix} , \quad \mathbf{w}_{y;k} \in \mathbb{R}^n \, . \tag{3.140}$$

By using $v_k$ from (3.96) and transformation in compact vector form, one obtains

$$\begin{aligned}\begin{pmatrix}\mathrm{E}[\mathbf{w}_{y;k}v_k]\\ 0\end{pmatrix} &= -\,\mathrm{E}[\begin{pmatrix}\hat{\hat{y}}^{(0)}_{m;k}\\ \vdots\\ \hat{\hat{y}}^{(n-1)}_{m;k}\\ 0\end{pmatrix}\begin{pmatrix}\hat{\hat{e}}^{(0)}_k & \cdots & \hat{\hat{e}}^{(n-1)}_k & \hat{\hat{e}}^{(n)}_k\end{pmatrix}\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\\ 1\end{pmatrix}]\\ &= -\,\mathrm{E}[\begin{pmatrix}\hat{\hat{y}}^{(0)}_{m;k}\hat{\hat{e}}^{(0)}_k & \cdots & \hat{\hat{y}}^{(0)}_{m;k}\hat{\hat{e}}^{(n)}_k\\ \vdots & \ddots & \vdots\\ \hat{\hat{y}}^{(n-1)}_{m;k}\hat{\hat{e}}^{(0)}_k & \cdots & \hat{\hat{y}}^{(n-1)}_{m;k}\hat{\hat{e}}^{(n)}_k\\ 0 & \cdots & 0\end{pmatrix}]\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\\ 1\end{pmatrix}\end{aligned} \tag{3.141}$$

where it is worth noting that the artificial extension by the $(n+1)$-th row is done for calculation reasons and is removed again. Replacing the noisy system output by its nominal value and additive random noise, i.e.,

$$y_{m;k} = y_k + e_k \tag{3.142}$$

results in

$$\begin{pmatrix}\mathrm{E}[\mathbf{w}_{y;k}v_k]\\ 0\end{pmatrix} = -\,\mathrm{E}[\begin{pmatrix}\hat{\hat{y}}^{(0)}_k\hat{\hat{e}}^{(0)}_k + \hat{\hat{e}}^{(0)}_k\hat{\hat{e}}^{(0)}_k & \cdots & \hat{\hat{y}}^{(0)}_k\hat{\hat{e}}^{(n)}_k + \hat{\hat{e}}^{(0)}_k\hat{\hat{e}}^{(n)}_k\\ \vdots & \ddots & \vdots\\ \hat{\hat{y}}^{(n-1)}_k\hat{\hat{e}}^{(0)}_k + \hat{\hat{e}}^{(n-1)}_k\hat{\hat{e}}^{(0)}_k & \cdots & \hat{\hat{y}}^{(n-1)}_k\hat{\hat{e}}^{(n)}_k + \hat{\hat{e}}^{(n-1)}_k\hat{\hat{e}}^{(n)}_k\\ 0 & \cdots & 0\end{pmatrix}]\begin{pmatrix}a_0\\ \vdots\\ a_{n-1}\\ 1\end{pmatrix}. \tag{3.143}$$

The $ij$-th element of the matrix in the above equation reads

$$\mathrm{E}[\mathbf{w}_{y;k}v_k]_{i,j} = -\left(\hat{\hat{y}}^{(i)}_k\,\mathrm{E}[\hat{\hat{e}}^{(j)}_k] + \mathrm{E}[\hat{\hat{e}}^{(i)}_k\hat{\hat{e}}^{(j)}_k]\right)\,,\quad 0 \le i \le n-1\,,\quad 0 \le j \le n\,. \tag{3.144}$$

Again, with $\mathrm{E}[e_k] = 0$, and the property of linear time-invariant (LTI) systems with random input, i.e.,

$$\mathrm{E}[\hat{\hat{e}}^{(j)}_k] = \mathrm{E}[e_k]\sum_{k=1}^{\infty} g^j_{F,k} \tag{3.145}$$

the first summand in (3.144) becomes zero. Hence, (3.143) simplifies to

$$\begin{pmatrix} \mathrm{E}[\mathbf{w}_{y;k} v_k] \\ 0 \end{pmatrix} = -\,\mathrm{E}[\begin{pmatrix} \hat{\hat{e}}_k^{(0)} \hat{\hat{e}}_k^{(0)} & \cdots & \hat{\hat{e}}_k^{(0)} \hat{\hat{e}}_k^{(n-1)} & \hat{\hat{e}}_k^{(0)} \hat{\hat{e}}_k^{(n)} \\ \vdots & \ddots & \vdots & \vdots \\ \hat{\hat{e}}_k^{(n-1)} \hat{\hat{e}}_k^{(0)} & \cdots & \hat{\hat{e}}_k^{(n-1)} \hat{\hat{e}}_k^{(n-1)} & \hat{\hat{e}}_k^{(n-1)} \hat{\hat{e}}_k^{(n)} \\ 0 & \cdots & 0 & 0 \end{pmatrix}] \begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \\ 1 \end{pmatrix} . \tag{3.146}$$

Introducing the vectors

$$\boldsymbol{\omega}^\top = \begin{pmatrix} \hat{\hat{e}}_k^{(0)} \\ \hat{\hat{e}}_k^{(1)} \\ \vdots \\ \hat{\hat{e}}_k^{(n-1)} \end{pmatrix} \in \mathbb{R}^n , \qquad \boldsymbol{\omega}_e^\top = \begin{pmatrix} \boldsymbol{\omega}^\top \\ \hat{\hat{e}}_k^{(n)} \end{pmatrix} = \begin{pmatrix} \hat{\hat{e}}_k^{(0)} \\ \hat{\hat{e}}_k^{(1)} \\ \vdots \\ \hat{\hat{e}}_k^{(n-1)} \\ \hat{\hat{e}}_k^{(n)} \end{pmatrix} \in \mathbb{R}^{n+1} \tag{3.147}$$

with the abbreviation for parameters related to the system output (and its time derivatives) $\mathbf{p}_a = (a_0, a_1, \cdots, a_{n-1}) \in \mathbb{R}^n$, and removing the last line, artificially added for calculation reasons, (3.146) gives

$$\mathrm{E}[\mathbf{w}_{y;k} v_k] = -\,\mathrm{E}[\boldsymbol{\omega}^\top \boldsymbol{\omega}_e] \begin{pmatrix} \mathbf{p}_a \\ 1 \end{pmatrix} = -\sigma^2 \boldsymbol{\Phi} \begin{pmatrix} \mathbf{p}_a \\ 1 \end{pmatrix} \tag{3.148}$$

where

$$\boldsymbol{\Phi} = \frac{1}{\sigma^2}\,\mathrm{E}[\boldsymbol{\omega}^\top \boldsymbol{\omega}_e] , \quad \boldsymbol{\Phi} \in \mathbb{R}^{n \times (n+1)} . \tag{3.149}$$

Using (3.135), the vectors $\boldsymbol{\omega}^\top$ and $\boldsymbol{\omega}_e^\top$ may be rewritten to

$$\boldsymbol{\omega}^\top = \begin{pmatrix} \hat{\hat{e}}_k^{(0)} \\ \hat{\hat{e}}_k^{(1)} \\ \vdots \\ \hat{\hat{e}}_k^{(n-1)} \end{pmatrix} = \begin{pmatrix} h_0^0 & h_1^0 & \cdots & h_n^0 \\ h_0^1 & h_1^1 & \cdots & h_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ h_0^{n-1} & h_1^{n-1} & \cdots & h_n^{n-1} \end{pmatrix} \begin{pmatrix} e_{F;k} \\ e_{F;k-1} \\ \vdots \\ e_{F;k-n} \end{pmatrix} = \boldsymbol{\kappa}^\top \mathbf{e}_F$$

$$\boldsymbol{\omega}_e^\top = \begin{pmatrix} \hat{\hat{e}}_k^{(0)} \\ \hat{\hat{e}}_k^{(1)} \\ \vdots \\ \hat{\hat{e}}_k^{(n-1)} \\ \hat{\hat{e}}_k^{(n)} \end{pmatrix} = \begin{pmatrix} h_0^0 & h_1^0 & \cdots & h_n^0 \\ h_0^1 & h_1^1 & \cdots & h_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ h_0^{n-1} & h_1^{n-1} & \cdots & h_n^{n-1} \\ h_0^n & h_1^n & \cdots & h_n^n \end{pmatrix} \begin{pmatrix} e_{F;k} \\ e_{F;k-1} \\ \vdots \\ e_{F;k-n} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\kappa}^\top \\ \hline \mathbf{h}^{n\top} \end{pmatrix} \mathbf{e}_F = \boldsymbol{\kappa}_e^\top \mathbf{e}_F \tag{3.150}$$

with filter coefficient matrices

$$\boldsymbol{\kappa}^\top = \begin{pmatrix} h_0^0 & h_1^0 & \cdots & h_n^0 \\ h_0^1 & h_1^1 & \cdots & h_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ h_0^{n-1} & h_1^{n-1} & \cdots & h_n^{n-1} \end{pmatrix} \in \mathbb{R}^{n\times(n+1)}\,, \qquad \boldsymbol{\kappa}_e^\top = \begin{pmatrix} \boldsymbol{\kappa}^\top \\ \hline \mathbf{h}^{n\top} \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} \tag{3.151}$$

and coefficient vector

$$\mathbf{h}^{n\top} = \begin{pmatrix} h_0^n & h_1^n & \cdots & h_n^n \end{pmatrix} \in \mathbb{R}^{n+1}\,. \tag{3.152}$$

The random noise $e_k$ is filtered with the common denominator

$$e_{F;k} = \frac{1}{d\left(z^{-1}\right)} e_k \tag{3.153}$$

where $\mathbf{e}_F$ contains the $k$-th element and $n$ causal shifts, i.e.,

$$\mathbf{e}_F = \begin{pmatrix} e_{F;k} \\ e_{F;k-1} \\ \vdots \\ e_{F;k-n} \end{pmatrix} \in \mathbb{R}^{n+1}\,. \tag{3.154}$$

Considering that the transpose of a product of matrices is the product of the transposes in the reverse order and taking into account the introduced abbreviations, one can rewrite (3.149) to

$$\begin{aligned} \boldsymbol{\Phi} &= \frac{1}{\sigma^2}\,\mathrm{E}[\boldsymbol{\kappa}^\top \mathbf{e}_F \mathbf{e}_F^\top \begin{pmatrix} \boldsymbol{\kappa} & \mathbf{h}^n \end{pmatrix}] = \frac{1}{\sigma^2}\boldsymbol{\kappa}^\top\,\mathrm{E}[\mathbf{e}_F \mathbf{e}_F^\top] \left(\boldsymbol{\kappa} \mid \mathbf{h}^n\right) \\ &= \frac{1}{\sigma^2}\boldsymbol{\kappa}^\top \boldsymbol{\Gamma} \boldsymbol{\kappa}_e \end{aligned} \tag{3.155}$$

where

$$\boldsymbol{\kappa}_e = \left(\boldsymbol{\kappa} \mid \mathbf{h}^n\right) \in \mathbb{R}^{(n+1)\times(n+1)}\,. \tag{3.156}$$

Matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{(n+1)\times(n+1)}$ denotes the autocorrelation function matrix of the filtered random noise $e_{F;k}$ and shows a unique structure called the Toeplitz matrix. The calculation rule reads

$$\boldsymbol{\Gamma}_{i,j} = \mathrm{E}[\mathbf{e}_F \mathbf{e}_F^\top]_{i,j} = r_{e_F e_F}\left(|i-j|\right) = r_{e_F e_F;|i-j|}\,, \quad 0 \le (i,j) \le n\,. \tag{3.157}$$

Again, let us assume $e_k$ to be a DT ergodic white noise process with independent samples. This implies

$$r_{ee;|i-j|} = \begin{cases} \sigma^2 , & |i-j| = 0 \\ 0 , & |i-j| \neq 0 \end{cases} \tag{3.158}$$

with constant finite noise variance $\sigma^2$. As shown in Section A.4.1, the autocorrelation function (ACF) of filtered white noise $r_{e_F e_F}$ is equivalent to the ACF of the impulse response $g_z$ of the corresponding filter, i.e.,

$$g_z(k) = \mathcal{Z}^{-1} \left\{ \frac{1}{d(z^{-1})} \right\} \tag{3.159}$$

multiplied by the noise variance $\sigma^2$. Consequently, (3.157) can be written to

$$\mathbf{\Gamma}_{i,j} = r_{e_F e_F;|i-j|} = \sigma^2 (g_z * g_z)(|i-j|) = \sigma^2 \sum_{k=-\infty}^{\infty} g_{z;k} g_{z;k-|i-j|} \,. \tag{3.160}$$

Finally, the asymptotic bias term from (3.126) can be expressed by

$$\Delta\mathbf{p} = N \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \, \mathrm{E}[\mathbf{w}_k v_k] = -N\sigma^2 \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \, \mathbf{\Phi} \begin{pmatrix} \mathbf{p}_a \\ 1 \end{pmatrix} \tag{3.161}$$

with

$$\mathbf{P} = \left( \mathbf{W}^\top \mathbf{W} \right)^{-1} \,. \tag{3.162}$$

**Asymptotically Unbiased Least-Squares Estimation using PMF Method**

**Result 3.3.** The asymptotically unbiased estimation result reads

$$\mathbf{p} = \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - \Delta\mathbf{p} = \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} + N\sigma^2 \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \, \mathbf{\Phi} \begin{pmatrix} \mathbf{p}_a \\ 1 \end{pmatrix} \tag{3.163}$$

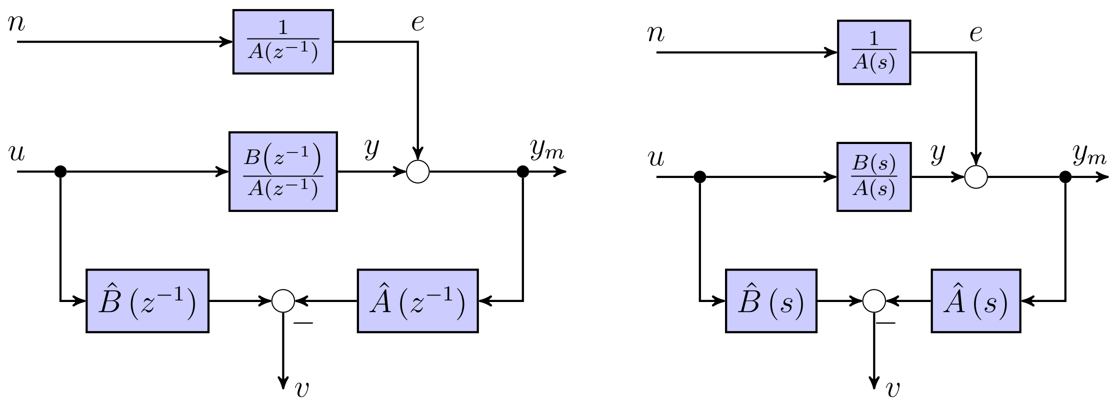
where $\hat{\mathbf{p}}$ and $\mathbf{P}$ are obtained by ordinary least-squares (OLS) estimation

$$\hat{\mathbf{p}} = \left( \mathbf{W}^\top \mathbf{W} \right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \tag{3.164}$$

and $\mathbf{p}_a = (a_0, a_1, \cdots, a_{n-1})^\top$ contains the nominal model parameters related to the system output, with noise variance $\sigma^2$.

#### 3.3.5.1 White Noise Filtered by Plant Dynamics

Literature research, e.g., [2, 3], reveals a well-known, artificially created framework for unbiased parameter estimation of time-discrete transfer functions. Namely, the filtering of white Gaussian noise with the denominator of the process to be identified, c.f., Fig. 3.3(a). Briefly summarized, the condition for consistency (3.125)




(a) Block diagram of a required structure for an unbiased estimation of DT transfer functions, where $n \sim i.i.d\,(0, \sigma^2)$, e.g., Gaussian white noise.

(b) Block diagram of a required structure for an unbiased estimation of DT transfer functions applied to CT transfer functions, where $n \sim i.i.d\,(0, \sigma^2)$.

Figure 3.3: Block diagram of a required structure for an unbiased estimation of DT transfer functions applied to CT transfer functions.

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\mathbf{W}^\top \mathbf{v} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_k v_k = \mathrm{E}[\mathbf{w}_k v_k] = 0 \tag{3.165}$$

is satisfied if the data vector $\mathbf{w}_k$ is uncorrelated with $v_k$. For DT models, the data vector

$$\mathbf{w}_k = \begin{pmatrix} -y_{m;k-1} \\ \vdots \\ -y_{m;k-n} \\ u_{k-1} \\ \vdots \\ u_{k-m_d} \end{pmatrix} \tag{3.166}$$

contains only past values of $y_m$ and $u$, i.e., causal shifts. Therefore, $\mathbf{w}_k$ is correlated with $\{v_{k-1}, v_{k-2}, \dots\}$, but uncorrelated with $v_k$. Consequently, condition (3.165) is fulfilled, and the parameter estimation is consistent, i.e., $\hat{\mathbf{p}} = \mathbf{p}$. Hence, $\hat{A}(s) = A(s)$

and $\hat{B}(s) = B(s)$, and consequently the composite noise term $v_k$ is also i.i.d., i.e.,

$$v(z) = n(z) \ . \tag{3.167}$$

For this reason, one may consider it important to apply this approach to CT systems. Assertion (3.167) is also valid in the CT case, i.e., $v(s) = n(s)$, but it is not helpful for the analysis since it requires unbiased estimation results.

The procedure for calculating the bias is entirely analogous to the calculation in Section 3.3.5, except that the additive output error is already obtained by convolution

$$e_k = (g_n * n)(k) \tag{3.168}$$

with impulse response $g_n = \mathcal{Z}^{-1}\left\{\frac{1}{A(z^{-1})}\right\}$. Hence, the PMF filtered error applies

$$\hat{\tilde{e}}_k^{(i)} = \left(g_F^i * e\right)(k) = \left(\left(g_n * g_F^i\right) * n\right)(k) = \left(g_{A,F}^i * n\right)(k) \tag{3.169}$$

with a modified filter

$$F_A^i\left(z^{-1}\right) = \frac{1}{A\left(z^{-1}\right)} F^i\left(z^{-1}\right) \tag{3.170}$$

and the corresponding impulse response

$$g_{A,F}^i(k) = \mathcal{Z}^{-1}\left\{F_A^i\left(z^{-1}\right)\right\} \ . \tag{3.171}$$

As shown in detail in Section 3.3.4, one obtains dependencies between regressor and noise vector with such a configuration, resulting in biased estimation results.

To summarize, contrary to DT models, the estimation result of CT systems is biased. Again, one can at least try to calculate the asymptotic bias to obtain asymptotically unbiased results. The calculation is analogous to Section 3.3.5, with the only difference that the measurement noise is convoluted with a modified filter.

## 3.4 Application to a Motivating Example

The task of parameter estimation and arising challenges, particularly the bias problem, are discussed in more detail during this section based on a 1D-servo positioning system with its nonlinear model equations

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \\ -\frac{d}{m}v - \frac{F_c}{m}\operatorname{sgn}(v) + \frac{1}{m}u \end{pmatrix} \tag{3.172}$$

with state vector $\mathbf{x} = (x, v)^\top \in \mathbb{R}^2$, parameter vector $\mathbf{p} = (m, d, F_\mathrm{c})^\top \in \mathbb{R}^3$, and input $u \in \mathbb{R}$. The system output $y$ is the servo position, i.e., $y = x \in \mathbb{R}$. Furthermore, it is assumed that the continuous-time output is measured at various discrete time instances $t_k = kT, 1 \leq k \leq N$ with sample time $T$. Taking this into account, (3.172) may be rewritten to

$$m\ddot{y}_k + d\dot{y}_k + F_\mathrm{c}\,\mathrm{sgn}\left(\dot{y}_k\right) = u_k \tag{3.173}$$

with nominal parameters

$$\mathbf{p} = (m, d, F_\mathrm{c})^\top \ . \tag{3.174}$$

Nonlinearities are not included in the introduced PMF-system class (c.f. (3.44)), especially concerning bias estimation. Nevertheless, it is possible to deal with particular nonlinearities, e.g., sign functions. The backward difference quotient approximates the velocity $v = \dot{y}$ in the sign function, and the sign function is approximated by a sum of positive and negative discretized unit steps, which only causes a small error as long as there is no chattering around zero velocity, e.g., [99]. For eliminating the time-derivatives in (3.173), the second-order stable linear filters

$$\begin{aligned} F^0(s) &= \frac{1}{\left(1 + sT_f\right)^2} \\ F^1(s) &= sF^0(s) \\ F^2(s) &= s^2F^0(s) = \frac{1}{T_f^2}\Big(1 - F^0(s) - 2T_fF^1(s)\Big) \end{aligned} \tag{3.175}$$

are introduced, where $T_f$ denotes the filter time constant.[10] In order to obtain only strictly proper transfer functions, partial fractional decomposition is applied to $F^2(s)$, whereby $F^2(s)$ can be replaced by a function of $F^0(s)$ and $F^1(s)$.

For application on sampled data, the filters have to be discretized. As described in Section 3.3.2, plant input filters and step function filters are discretized in an exact manner by zero-order-hold (ZOH), filters applied to the continuous, but sampled, plant output $y_k$ are discretized by bilinear approximation. Discretization of $F^0(s)$ by ZOH results in

$$F^0_{zoh}(z) = \frac{e^{\frac{T}{T_f}}T + T_f - e^{\frac{T}{T_f}}T_f + z\left(-e^{\frac{T}{T_f}}T - e^{\frac{T}{T_f}}T_f + e^{\frac{2T}{T_f}}T_f\right)}{T_f - z2e^{\frac{T}{T_f}}T_f + z^2e^{\frac{2T}{T_f}}T_f} \tag{3.176}$$

[10]The filter time constant $T_f$ must be selected sensibly according to the application.

with the corresponding impulse response

$$g^0_{m,zoh}(k) = \mathcal{Z}^{-1}\{F^0_{zoh}(z)\} = \frac{1}{T_f}\left(e^{-\frac{kT}{T_f}}\left(-kT - T_f + e^{\frac{T}{T_f}}\left((k-1)T + T_f\right)\right)\right) . \tag{3.177}$$

Applying bilinear approximation (3.58) to $F^0(s)$ and $F^1(s)$, respectively, and calculating the respective impulse response sequences by application of the inverse z-transform, cf. (3.59), results in

$$\begin{aligned} F^0_{tust}(z) &= \frac{T^2 + z2T^2 + z^2T^2}{T^2 - 4TT_f + 4T_f^2 + z\left(2T^2 - 8T_f^2\right) + z^2\left(T^2 + 4TT_f + 4T_f^2\right)} \\ F^1_{tust}(z) &= \frac{-2T + z^2 2T}{T^2 - 4TT_f + 4T_f^2 + z\left(2T^2 - 8T_f^2\right) + z^2\left(T^2 + 4TT_f + 4T_f^2\right)} \end{aligned} \tag{3.178}$$

and

$$\begin{aligned} g^0_{m,tust}(k) &= \mathcal{Z}^{-1}\{F^0_{tust}(z)\} = -\frac{8T^2T_f\left(T - 2kT_f\right)\left(1 - \frac{2T}{T+2T_f}\right)^k}{\left(T^2 - 4T_f^2\right)^2} \\ g^1_{m,tust}(k) &= \mathcal{Z}^{-1}\{F^1_{tust}(z)\} = \frac{4T\left(T^2 - 4kTT_f + 4T_f^2\right)\left(1 - \frac{2T}{T+2T_f}\right)^k}{\left(T^2 - 4T_f^2\right)^2} . \end{aligned} \tag{3.179}$$

Applying PMF filtering to (3.173) results in the equation for identification for one single measurement

$$\underbrace{\tilde{u}^{(0)}_k}_{=y_{LSQ;k}} = \underbrace{\begin{pmatrix} \frac{1}{T_f^2}\left(y_k - \hat{\tilde{y}}^{(0)}_k - 2T_f\hat{\tilde{y}}^{(1)}_k\right) & \hat{\tilde{y}}^{(1)}_k & \widetilde{\operatorname{sgn}(\dot{y})}^{(0)}_k \end{pmatrix}}_{=\mathbf{w}_k^\top} \underbrace{\begin{pmatrix} m \\ d \\ F_c \end{pmatrix}}_{=\mathbf{p}} \tag{3.180}$$

or in compact vector form

$$y_{LSQ;k} = \mathbf{w}_k^\top \mathbf{p} \tag{3.181}$$

with the filtered signals

$$\begin{aligned} \widetilde{\operatorname{sgn}(\dot{y})}^{(0)}_k &= \left(g^0_{m,zoh} * \operatorname{sgn}(\dot{y})\right)(k) & \hat{\tilde{y}}^{(0)}_k &= \left(g^0_{m,tust} * y\right)(k) \\ \tilde{u}^{(0)}_k &= \left(g^0_{m,zoh} * u\right)(k) & \hat{\tilde{y}}^{(1)}_k &= \left(g^1_{m,tust} * y\right)(k) . \end{aligned} \tag{3.182}$$

Using $N$, $N > n_p$ samples, (3.181) results in a system of algebraic equations which has the same structure as (3.68), i.e.,

$$\mathbf{y}_{LSQ} = \mathbf{W}\mathbf{p} \tag{3.183}$$

with $\mathbf{y}_{LSQ} \in \mathbb{R}^N$, $\mathbf{W} \in \mathbb{R}^{N\times 3}$ and $\mathbf{p} \in \mathbb{R}^3$, whose unknown parameters can be estimated by solving

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \,. \tag{3.184}$$

### 3.4.1 Analytical Solution for the Estimation Bias Caused by PMF Approach

The measured samples are affected by an additive random noise $e_k$, i.e.,

$$y_{m;k} = y_m(t_k) = y_k + e_k, \quad 1 \le k \le N\,. \tag{3.185}$$

One assumes that the random variable $e_k$ is a realization of an ergodic, i.e., stationary, white noise process with its expected value $\mathrm{E}[e_k] = 0$ and constant variance $\sigma^2$. Again, retaining the measured samples, i.e., $y = y_m - e$, the equation for identification reads[11]

$$\begin{aligned}\underbrace{\tilde{u}_k^{(0)}}_{=y_{LSQ;k}} &= \underbrace{\begin{pmatrix}\frac{1}{T_f^2}\left(y_{m;k} - \hat{\tilde{y}}_{m;k}^{(0)} - 2T_f \hat{\tilde{y}}_{m;k}^{(1)}\right) & \hat{\tilde{y}}_{m;k}^{(1)} & \widetilde{\mathrm{sgn}\left(\dot{y}_m - \dot{e}\right)}_k^{(0)}\end{pmatrix}}_{=\mathbf{w}_k^\top} \underbrace{\begin{pmatrix} m \\ d \\ F_\mathrm{c}\end{pmatrix}}_{=\mathbf{p}} \\ &\quad + \underbrace{(-1)\begin{pmatrix}\frac{1}{T_f^2}\left(e_k - \hat{\tilde{e}}_k^{(0)} - 2T_f \hat{\tilde{e}}_k^{(1)}\right) & \hat{\tilde{e}}_k^{(1)}\end{pmatrix}\begin{pmatrix} m \\ d\end{pmatrix}}_{=v_k}\end{aligned} \tag{3.186}$$

where

$$\begin{aligned}\hat{\tilde{e}}_k^{(0)} &= \left(g^0_{m,tust} * e\right)(k) \\ \hat{\tilde{e}}_k^{(1)} &= \left(g^1_{m,tust} * e\right)(k) \;.\end{aligned} \tag{3.187}$$

As shown in Section 3.3.4, the main task in calculating the bias is to derive an analytical expression of $\mathrm{E}[\mathbf{w}_k v_k]$, which is only determined by the chosen signal filters $F^0(s)$, $F^1(s)$, or the discretized filters $F^0_{zoh}(z)$, $F^0_{tust}(z)$ and $F^1_{tust}(z)$, respectively. One can

[11]Please note that in contrast to the chosen normalization in (3.96), $y_{LSQ}$ is not the highest derivative of the output. The reason for this is that the system parameters thus appear linear and separated in the equation. In principle, the standardization can be chosen freely. However, it is important to note that the result from (3.163) cannot be adopted one-to-one.

separate the transposed data vector $\mathbf{w}_k$ into two parts

$$\mathbf{w}_k = \begin{pmatrix} \mathbf{w}_{y;k} \\ w_{c;k} \end{pmatrix} , \quad \mathbf{w}_k \in \mathbb{R}^3 \tag{3.188}$$

where

$$\mathbf{w}_{y;k} = \begin{pmatrix} \frac{1}{T_f^2} \left( y_{m;k} - \hat{\tilde{y}}_{m;k}^{(0)} - 2T_f \hat{\tilde{y}}_{m;k}^{(1)} \right) & \hat{\tilde{y}}_{m;k}^{(1)} \end{pmatrix}^\top , \quad \mathbf{w}_{y;k} \in \mathbb{R}^2 \tag{3.189}$$

is directly affected by the random output noise and

$$w_{c;k} = \widetilde{\mathrm{sgn}\,(\dot{y})}_k^{(0)} = \widetilde{\mathrm{sgn}\,(\dot{y}_m - \dot{e})}_k^{(0)} , \quad w_{c;k} \in \mathbb{R} \tag{3.190}$$

is a unknown but deterministic term due to

$$\mathrm{sgn}\,(\dot{y}_m - \dot{e})_k = \begin{cases} -1 , & \dot{y}_{m;k} - \dot{e}_k < 0 \\ 0 , & \dot{y}_{m;k} - \dot{e}_k = 0 \\ +1 , & \dot{y}_{m;k} - \dot{e}_k > 0 \end{cases} . \tag{3.191}$$

Inserting (3.188) in $\mathrm{E}[\mathbf{w}_k v_k]$ results in

$$\mathrm{E}[\mathbf{w}_k v_k] = \mathrm{E}[ \begin{pmatrix} \mathbf{w}_{y;k} \\ w_{c;k} \end{pmatrix} v_k] = \mathrm{E}[ \begin{pmatrix} \mathbf{w}_{y;k} v_k \\ w_{c;k} v_k \end{pmatrix} ] = \begin{pmatrix} \mathrm{E}[\mathbf{w}_{y;k} v_k] \\ \mathrm{E}[w_{c;k} v_k] \end{pmatrix} . \tag{3.192}$$

Consider stable DT LTI systems, whose impulse responses are $g_{m,tust}^{(0)}$ and $g_{m,tust}^{(1)}$ and whose input is a WSS random process $e_k$. In [115, 116], the authors show that the output processes $\hat{\tilde{e}}_k^{(0)}$ and $\hat{\tilde{e}}_k^{(1)}$ are also WSS, and it applies

$$\begin{aligned} \mathrm{E}[\hat{\tilde{e}}_k^{(0)}] &= \mathrm{E}[e_k] \sum_{m=-\infty}^{\infty} g_{F,tust;m}^{0} \\ \mathrm{E}[\hat{\tilde{e}}_k^{(1)}] &= \mathrm{E}[e_k] \sum_{m=-\infty}^{\infty} g_{F,tust;m}^{1} . \end{aligned} \tag{3.193}$$

With the noise assumption

$$\mathrm{E}[e_k] = 0 \tag{3.194}$$

one gets

$$\begin{aligned} \mathrm{E}[\hat{\tilde{e}}_k^{(0)}] &= 0 \\ \mathrm{E}[\hat{\tilde{e}}_k^{(1)}] &= 0 . \end{aligned} \tag{3.195}$$

Consequently, for the second term, one obtains

$$
\begin{aligned}
\mathrm{E}[w_{c;k}v_k] &= -\,\mathrm{E}[\operatorname{sgn}\widetilde{(\dot{y}_m - \dot{e})}_k^{(0)} \begin{pmatrix} \frac{1}{T_f^2}\left(e_k - \hat{\hat{e}}_k^{(0)} - 2T_f\hat{\hat{e}}_k^{(1)}\right) & \hat{\hat{e}}_k^{(1)} \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix}] \\
&= -\operatorname{sgn}\widetilde{(\dot{y}_m - \dot{e})}_k^{(0)}\, \mathrm{E}[\begin{pmatrix} \frac{1}{T_f^2}\left(e_k - \hat{\hat{e}}_k^{(0)} - 2T_f\hat{\hat{e}}_k^{(1)}\right) & \hat{\hat{e}}_k^{(1)} \end{pmatrix}] \begin{pmatrix} m \\ d \end{pmatrix} \\
&= -\operatorname{sgn}\widetilde{(\dot{y}_m - \dot{e})}_k^{(0)} \begin{pmatrix} \frac{1}{T_f^2}\left(\mathrm{E}[e_k] - \mathrm{E}[\hat{\hat{e}}_k^{(0)}] - 2T_f\,\mathrm{E}[\hat{\hat{e}}_k^{(1)}]\right) & \mathrm{E}[\hat{\hat{e}}_k^{(1)}] \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix} = 0\,.
\end{aligned}
\tag{3.196}
$$

For better readability, the exact expressions

$$
\begin{aligned}
\hat{\hat{y}}_{m;k}^{(2)} &= \frac{1}{T_f^2}\left(y_{m;k} - \hat{\hat{y}}_{m;k}^{(0)} - 2T_f\hat{\hat{y}}_{m;k}^{(1)}\right) \\
\hat{\hat{e}}_k^{(2)} &= \frac{1}{T_f^2}\left(e_k - \hat{\hat{e}}_k^{(0)} - 2T_f\hat{\hat{e}}_k^{(1)}\right)
\end{aligned}
\tag{3.197}
$$

are introduced as abbreviations. Analogous to (3.193), we can write

$$
\mathrm{E}[\hat{\hat{e}}_k^{(2)}] = \frac{1}{T_f^2}\left(\mathrm{E}[e_k] - \mathrm{E}[\hat{\hat{e}}_k^{(0)}] - 2T_f\,\mathrm{E}[\hat{\hat{e}}_k^{(1)}]\right) = 0\,. \tag{3.198}
$$

Replacing the noise-affected system output $y_{m;k}$ by its nominal value and additive random noise, c.f. (3.185), moreover, considering (3.195) and (3.198), respectively, one obtains

$$
\begin{aligned}
\mathrm{E}[\mathbf{w}_{y;k}v_k] &= -\,\mathrm{E}[\begin{pmatrix} \hat{\hat{y}}_k^{(2)} + \hat{\hat{e}}_k^{(2)} \\ \hat{\hat{y}}_k^{(1)} + \hat{\hat{e}}_k^{(1)} \end{pmatrix} \begin{pmatrix} \hat{\hat{e}}_k^{(2)} & \hat{\hat{e}}_k^{(1)} \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix}] \\
&= -\,\mathrm{E}[\begin{pmatrix} \hat{\hat{y}}_k^{(2)}\hat{\hat{e}}_k^{(2)} + \hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(2)} & \hat{\hat{y}}_k^{(2)}\hat{\hat{e}}_k^{(1)} + \hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(1)} \\ \hat{\hat{y}}_k^{(1)}\hat{\hat{e}}_k^{(2)} + \hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(2)} & \hat{\hat{y}}_k^{(1)}\hat{\hat{e}}_k^{(1)} + \hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(1)} \end{pmatrix}] \begin{pmatrix} m \\ d \end{pmatrix} \\
&= -\begin{pmatrix} \hat{\hat{y}}_k^{(2)}\,\mathrm{E}[\hat{\hat{e}}_k^{(2)}] + \mathrm{E}[\hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(2)}] & \hat{\hat{y}}_k^{(2)}\,\mathrm{E}[\hat{\hat{e}}_k^{(1)}] + \mathrm{E}[\hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(1)}] \\ \hat{\hat{y}}_k^{(1)}\,\mathrm{E}[\hat{\hat{e}}_k^{(2)}] + \mathrm{E}[\hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(2)}] & \hat{\hat{y}}_k^{(1)}\,\mathrm{E}[\hat{\hat{e}}_k^{(1)}] + \mathrm{E}[\hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(1)}] \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix} \\
&= -\begin{pmatrix} \mathrm{E}[\hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(2)}] & \mathrm{E}[\hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(1)}] \\ \mathrm{E}[\hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(2)}] & \mathrm{E}[\hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(1)}] \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix} \\
&= -\,\mathrm{E}[\begin{pmatrix} \hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(2)} & \hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(1)} \\ \hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(2)} & \hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(1)} \end{pmatrix}] \begin{pmatrix} m \\ d \end{pmatrix} \\
&= -\sigma^2\boldsymbol{\Phi} \begin{pmatrix} m \\ d \end{pmatrix}
\end{aligned}
\tag{3.199}
$$

with

$$\begin{aligned}\mathbf{\Phi} &= \frac{1}{\sigma^2}\,\mathrm{E}[\begin{pmatrix}\hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(2)} & \hat{\hat{e}}_k^{(2)}\hat{\hat{e}}_k^{(1)}\\ \hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(2)} & \hat{\hat{e}}_k^{(1)}\hat{\hat{e}}_k^{(1)}\end{pmatrix}]\\ &= \frac{1}{\sigma^2}\,\mathrm{E}[\tilde{\boldsymbol{\omega}}^\top\tilde{\boldsymbol{\omega}}]\end{aligned} \tag{3.200}$$

where $\mathbf{\Phi} \in \mathbb{R}^{2\times 2}$ and

$$\boldsymbol{\omega}^\top = \begin{pmatrix}\hat{\hat{e}}_k^{(2)}\\ \hat{\hat{e}}_k^{(1)}\end{pmatrix} = \begin{pmatrix}\frac{1}{T_f^2}\left(e_k - \hat{\hat{e}}_k^{(0)} - 2T_f\hat{\hat{e}}_k^{(1)}\right)\\ \hat{\hat{e}}_k^{(1)}\end{pmatrix}, \quad \boldsymbol{\omega}^\top \in \mathbb{R}^2\,. \tag{3.201}$$

Dividing the discrete filters $F^0_{tust}(z)$ and $F^1_{tust}(z)$ (3.178) by $z^2$ results in causal IIR filters[12]

$$F^0_{tust}\left(z^{-1}\right) = \frac{\sum_{i=0}^2 h_i^{(0)} z^{-i}}{d\left(z^{-1}\right)} \quad \text{and} \quad F^i_{tust}\left(z^{-1}\right) = \frac{\sum_{i=0}^2 h_i^{(1)} z^{-i}}{d\left(z^{-1}\right)} \tag{3.202}$$

with the common denominator

$$d\left(z^{-1}\right) = 1 + \underbrace{\frac{2T^2 - 8T_f^2}{T^2 + 4TT_f + 4T_f^2}}_{=d_1} z^{-1} + \underbrace{\frac{T^2 - 4TT_f + 4T_f^2}{T^2 + 4TT_f + 4T_f^2}}_{=d_2} z^{-2} \tag{3.203}$$

with denominator coefficients $d_1$, $d_2$ and the respective numerator coefficients

$$\begin{aligned}h_0^{(0)} &= \frac{T^2}{(T+2T_f)^2} & h_1^{(0)} &= \frac{2T^2}{(T+2T_f)^2} & h_2^{(0)} &= \frac{T^2}{(T+2T_f)^2}\\ h_0^{(1)} &= \frac{2T}{(T+2T_f)^2} & h_1^{(1)} &= 0 & h_2^{(1)} &= -\frac{2T}{(T+2T_f)^2}\,.\end{aligned} \tag{3.204}$$

Please note that $z^{-1}$ is just the inverted from $z$. The reason for this transformation is that the inverse z transform of the occurring signals only contains causal shifts (lag operator), resulting in, e.g., $z^{-1}e_k = e_{k-1}$. Again, to improve readability, one may introduce the exact expressions

$$\begin{aligned}h_0^{(2)} &= \left(\frac{1}{T_f^2}\left(1 - h_0^{(0)} - 2T_f h_0^{(1)}\right)\right)\\ h_1^{(2)} &= \left(\frac{1}{T_f^2}\left(d_1 - h_1^{(0)} - 2T_f h_1^{(1)}\right)\right)\\ h_2^{(2)} &= \left(\frac{1}{T_f^2}\left(d_2 - h_2^{(0)} - 2T_f h_2^{(1)}\right)\right)\,.\end{aligned} \tag{3.205}$$

[12] Please note that the notation is changed from $F^i_{tust}(z)$ to $F^i_{tust}\left(z^{-1}\right)$ for computing $\tilde{\boldsymbol{\omega}}$.

Then, (3.201) may be rewritten using (3.202) and (3.205) in vector form to

$$\begin{aligned}\tilde{\boldsymbol{\omega}}^\top &= \begin{pmatrix} h_0^{(2)} e_{F;k} + h_1^{(2)} e_{F;k-1} + h_2^{(2)} e_{F;k-2} \\ h_0^{(1)} e_{F;k} + h_2^{(1)} e_{F;k-2} \end{pmatrix} = \underbrace{\begin{pmatrix} h_0^{(2)} & h_1^{(2)} & h_2^{(2)} \\ h_0^{(1)} & h_1^{(1)} & h_2^{(1)} \end{pmatrix}}_{=\boldsymbol{\kappa}^\top} \underbrace{\begin{pmatrix} e_{F;k} \\ e_{F;k-1} \\ e_{F;k-2} \end{pmatrix}}_{=\mathbf{e}_F} \\ &= \boldsymbol{\kappa}^\top \mathbf{e}_F\end{aligned} \tag{3.206}$$

with

$$\mathbf{e}_F = \begin{pmatrix} e_{F;k} \\ e_{F;k-1} \\ e_{F;k-2} \end{pmatrix} \in \mathbb{R}^3 \tag{3.207}$$

where

$$e_{F;k} = \frac{1}{d\left(z^{-1}\right)} e_k = \frac{\left(T + 2T_f\right)^2 z^2}{\left(2T_f\left(z-1\right) + T\left(z+1\right)\right)^2} e_k \tag{3.208}$$

and the filter coefficient matrix

$$\boldsymbol{\kappa} = \begin{pmatrix} h_0^{(2)} & h_0^{(1)} \\ h_1^{(2)} & h_1^{(1)} \\ h_2^2 & h_2^{(1)} \end{pmatrix} \in \mathbb{R}^{3\times 2}\,. \tag{3.209}$$

Equation (3.200) can be expressed as more compact

$$\begin{aligned}\boldsymbol{\Phi} &= \frac{1}{\sigma^2}\,\mathrm{E}[\tilde{\boldsymbol{\omega}}^\top \tilde{\boldsymbol{\omega}}] = \frac{1}{\sigma^2}\,\mathrm{E}[\boldsymbol{\kappa}^\top \mathbf{e}_F \mathbf{e}_F^\top \boldsymbol{\kappa}] = \frac{1}{\sigma^2} \boldsymbol{\kappa}^\top\,\mathrm{E}[\mathbf{e}_F \mathbf{e}_F^\top] \boldsymbol{\kappa} \\ &= \frac{1}{\sigma^2} \boldsymbol{\kappa}^\top \boldsymbol{\Gamma} \boldsymbol{\kappa}\end{aligned} \tag{3.210}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{3\times 3}$ has a special structure called the Toeplitz matrix. Matrix $\boldsymbol{\Gamma}$ denotes the expectation of the sample covariance matrix of the filtered random noise sequence $e_{F;k}$ with a single matrix entry ($i$-th row, $j$-th column)

$$\boldsymbol{\Gamma}_{i,j} = \mathrm{E}[e_{F;k-i} e_{F;k-j}]\,, \quad 0 \le (i,j) \le 2\,. \tag{3.211}$$

Per definition, the autocorrelation function of a WSS process is equal to the expected value of the random variable or signal realization with a time-shifted version of itself, e.g., [117]. Using (A.61), the expectation value in (3.211) can be replaced by

$$\boldsymbol{\Gamma}_{i,j} = r_{e_F e_F}\left(|i-j|\right) = r_{e_F e_F;|i-j|}\,, \quad 0 \le (i,j) \le 2 \tag{3.212}$$

where $r_{e_F e_F}$ is the autocorrelation function of the filtered random noise process $e_F$. Assuming that the random variables of the random process $e$ are uncorrelated in time (serially uncorrelated)

$$r_{ee;|i-j|} = \begin{cases} \sigma^2 & |i-j| = 0 \\ 0 & |i-j| \neq 0 \end{cases} \tag{3.213}$$

with finite constant variance $\sigma^2$, we call the random process a *white noise process*, see (A.83). The autocorrelation of filtered white noise is equal to the filter's impulse response times the noise variance (see Section A.4.1). By this property (3.212) can be rewritten to

$$\mathbf{\Gamma}_{i,j} = r_{e_F e_F;|i-j|} = \sigma^2 \left(g_z * g_z\right)\left(|i-j|\right) = \sigma^2 \sum_{k=-\infty}^{\infty} g_{z;k} g_{z;k-|i-j|} \tag{3.214}$$

with impulse response

$$g_z\left(k\right) = \mathcal{Z}^{-1}\left\{\frac{1}{d\left(z^{-1}\right)}\right\} = (1+k)\left(-\frac{T-2T_f}{T+2T_f}\right)^k . \tag{3.215}$$

Building $\mathbf{\Gamma}$ and inserting it in (3.210) results in

$$\mathbf{\Phi} = \begin{pmatrix} \frac{T+4T_f}{T_f^3\left(T+2T_f\right)^2} & 0 \\ 0 & \frac{T}{T_f\left(T+2T_f\right)^2} \end{pmatrix} , \quad \mathbf{\Phi} \in \mathbb{R}^{2\times 2} . \tag{3.216}$$

Joining (3.188), (3.196), (3.199), and (3.216) gives

$$\mathrm{E}[\mathbf{w}_k v_k] = -\sigma^2 \begin{pmatrix} \begin{pmatrix} \frac{T+4T_f}{T_f^3\left(T+2T_f\right)^2} & 0 \\ 0 & \frac{T}{T_f\left(T+2T_f\right)^2} \end{pmatrix} \begin{pmatrix} m \\ d \end{pmatrix} \\ 0 \end{pmatrix} = -\sigma^2 \begin{pmatrix} \frac{T+4T_f}{T_f^3\left(T+2T_f\right)^2} m \\ \frac{T}{T_f\left(T+2T_f\right)^2} d \\ 0 \end{pmatrix} \tag{3.217}$$

or more compact

$$\mathrm{E}[\mathbf{w}_k v_k] = -\sigma^2 \mathbf{\Phi}_e \mathbf{p} \tag{3.218}$$

where $\mathbf{\Phi}_e \in \mathbb{R}^{3\times 3}$ denotes

$$\mathbf{\Phi}_e = \begin{pmatrix} \mathbf{\Phi} & \mathbf{0}_{(2\times 1)} \\ \mathbf{0}_{(1\times 2)} & 0 \end{pmatrix} . \tag{3.219}$$

The asymptotic bias term $\Delta\mathbf{p}$[13] is

$$\Delta\mathbf{p} = -\sigma^2 N \operatorname*{plim}_{N\to\infty} (\mathbf{P})\, \mathbf{\Phi}_e \mathbf{p} = -\sigma^2 N \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \begin{pmatrix} \frac{T+4T_f}{T_f^3\left(T+2T_f\right)^2} m \\ \frac{T}{T_f\left(T+2T_f\right)^2} d \\ 0 \end{pmatrix} \tag{3.220}$$

with

$$\mathbf{P} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1}, \quad \mathbf{P} \in \mathbb{R}^{3\times 3}. \tag{3.221}$$

**Asymptotically Unbiased OLS Estimation for 1D-Positioning System**

**Result 3.4.** The asymptotically unbiased least-squares estimation for offline evaluation is finally obtained by

$$\begin{aligned} \mathbf{p} &= \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - \Delta\mathbf{p} = \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} + \sigma^2 N \operatorname*{plim}_{N\to\infty} (\mathbf{P})\, \mathbf{\Phi}_e \mathbf{p} \\ &= \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} + \sigma^2 N \operatorname*{plim}_{N\to\infty} (\mathbf{P}) \begin{pmatrix} \frac{T+4T_f}{T_f^3\left(T+2T_f\right)^2} m \\ \frac{T}{T_f\left(T+2T_f\right)^2} d \\ 0 \end{pmatrix} \end{aligned} \tag{3.222}$$

whereby the nominal model parameters $(m,d)$ and the noise variance $\sigma^2$ must be known.

### 3.4.2 Implementation Aspects

For implementation purposes, it is worth mentioning that the nominal model parameters and the variance are unknown and need to be replaced by estimates, so $\Delta\mathbf{p} \to \Delta\hat{\mathbf{p}}$, $\mathbf{p} \to \hat{\mathbf{p}}_{BC}$, and $\sigma \to \hat{\sigma}$.

[13] Please note that the calculation of the asymptotic bias term is based on approximately discretized signal filters. For these signal filters, the calculated bias term is exact, not an estimate.

**Implementation: Unbiased OLS Estimation for 1D-Positioning System**

**Result 3.5.** Equation (3.222) changes slightly to

$$\hat{\mathbf{p}}_{\mathrm{BC}} = \hat{\mathbf{p}} - \Delta\hat{\mathbf{p}} = \hat{\mathbf{p}} + \hat{\sigma}^2 N \operatorname*{plim}_{N\to\infty}(\mathbf{P})\,\mathbf{\Phi}_e\hat{\mathbf{p}} = \hat{\mathbf{p}} + \hat{\sigma}^2 N \operatorname*{plim}_{N\to\infty}(\mathbf{P}) \begin{pmatrix} \frac{T+4T_f}{T_f^3\left(T+2T_f\right)^2}\hat{m} \\ \frac{T}{T_f\left(T+2T_f\right)^2}\hat{d} \\ 0 \end{pmatrix}. \tag{3.223}$$

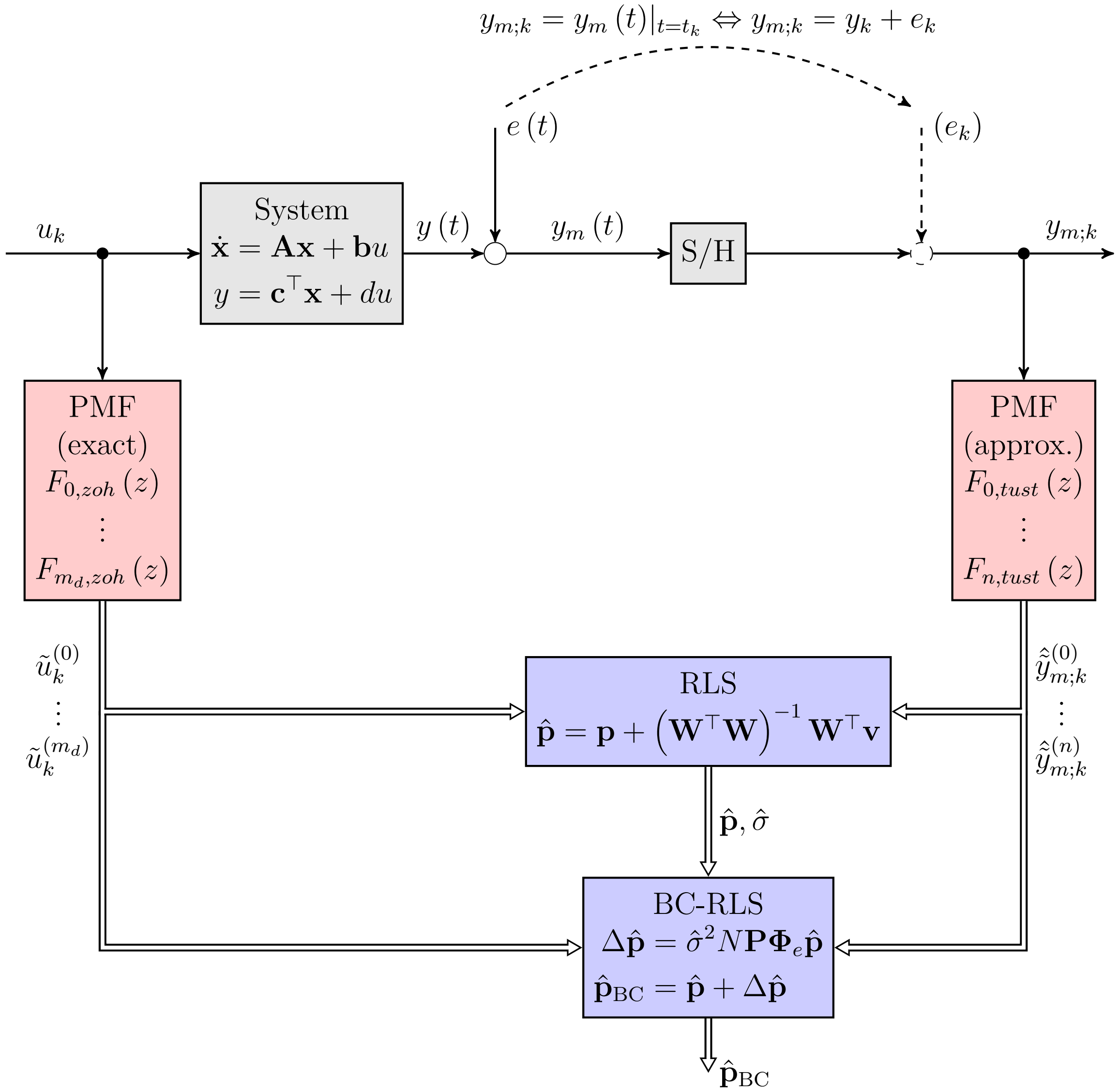



Figure 3.4: Block diagram of the proposed bias-compensating recursive least-squares algorithm.

#### 3.4.2.1 Recursive Algorithm and Noise Variance Estimation

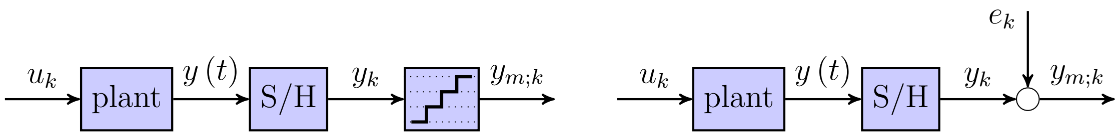
In [7], a recursive implementation of the derived bias-compensating algorithm is presented, suitable for online parameter estimation. Figure 3.4 shows the block diagram of the proposed online capable algorithm. Additionally, the proposed algorithm includes the estimation of the unknown noise variance. It reads

$$\begin{aligned} e_{LSQ,k} &= y_{LSQ;k} - \mathbf{w}_k^\top \hat{\mathbf{p}}_{k-1} \\ \mathbf{L}_k &= \frac{\mathbf{P}_{k-1}\mathbf{w}_k}{\lambda + \mathbf{w}_k^\top \mathbf{P}_{k-1}\mathbf{w}_k} \\ \hat{\mathbf{p}}_k &= \hat{\mathbf{p}}_{k-1} + \mathbf{L}_k e_{LSQ,k} \\ \mathbf{P}_k &= \frac{1}{\lambda}\left(\mathbf{P}_{k-1} - \mathbf{L}_k \mathbf{w}_k^\top \mathbf{P}_{k-1}\right) \\ R_k &= R_{k-1} + \frac{e_{LSQ,k}^2}{1 + \mathbf{w}_k^\top \mathbf{P}_{k-1}\mathbf{w}_k} \\ \hat{c}_k &= \hat{\mathbf{p}}_k^\top \mathbf{\Phi}_e \hat{\mathbf{p}}_{\mathrm{BC},k-1} \\ \hat{\mathbf{p}}_{\mathrm{BC},k} &= \hat{\mathbf{p}}_k + \mathbf{P}_k \frac{R_k}{\hat{c}_k} \mathbf{\Phi}_e \hat{\mathbf{p}}_{\mathrm{BC},k-1} \end{aligned} \tag{3.224}$$

where $\lambda$, $0 < \lambda \leq 1$ describes the well-known exponential forgetting factor.

### 3.4.3 Simulation Results

Assuming significant changes from sample to sample, it is common practice to approximate quantized system outputs by its pure sampled output and an additive sequence of uncorrelated white noise with constant and known variance, e.g., [99, 122–124]. It



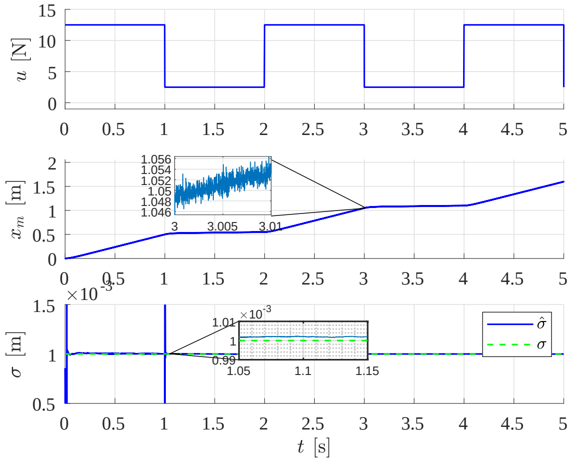

(a) Block diagram of quantization error. (b) Approximation.

Figure 3.5: Modeling quantization error caused by, e.g., incremental encoders, as additive white noise.

follows directly that the additive white noise approach commonly used in system identification is highly relevant in practice. For example, it can be used to model sensor noise, e.g., quantization effects of incremental encoders, c.f. Fig. 3.5.

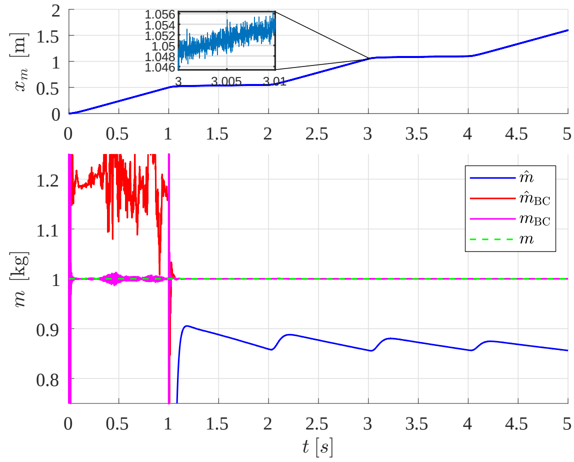
A comparison between standard recursive least-squares (RLS) and the proposed bias-compensated recursive least-squares (BC-RLS) algorithm is carried out in simulation.

The sample time is set to $T = 10\,\mu\mathrm{s}$, whereas the time constant of the PMF filters is set to $T_f = 50\,\mathrm{ms}$. The "measured" output $x_m$ is noise corrupted, where the noise $e$ is white, and Gaussian distributed, with zero mean and standard deviation $\sigma_x = 0.001\,\mathrm{m}$. The nominal model parameters read: $m = 1\,\mathrm{kg}$, $d = 20\,\mathrm{N\,s\,m^{-1}}$, and $F_c = 2\,\mathrm{N}$.

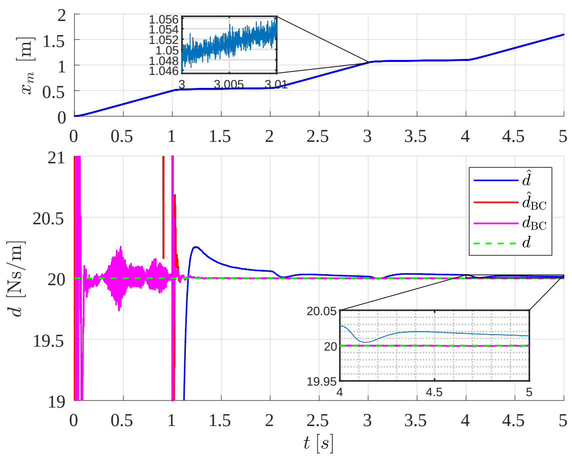



(a) Cart input force $u$, noise affected system output $x_m$ and nominal and (estimated) noise std. deviation $\hat{\sigma}_x$

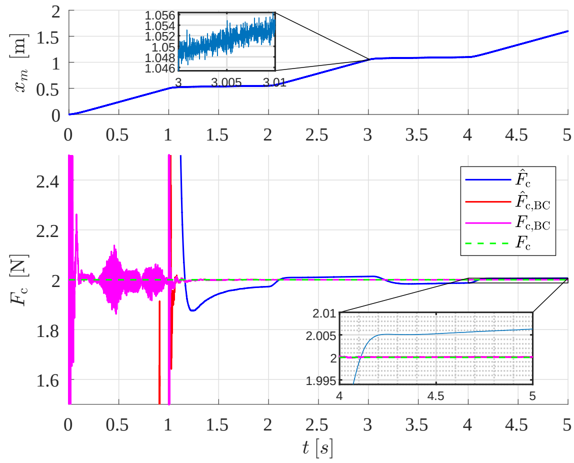
.




(b) Measured output $x_m$, biased OLS estimate $\hat{m}$, bias-compensated estimated values $\hat{m}_{\mathrm{BC}}$ and $m_{\mathrm{BC}}$ (nom. par. known), and nominal value $m = 1\,\mathrm{kg}$.




(c) Measured output $x_m$, biased OLS estimate $\hat{d}$, bias-compensated estimated values $\hat{d}_{\mathrm{BC}}$ and $d_{\mathrm{BC}}$ (nom. par. known), and nominal value $d = 20\,\mathrm{N\,s\,m^{-1}}$.




(d) Measured output $x_m$, biased OLS estimate $\hat{F}_c$, bias-compensated estimated values $\hat{F}_{c,\mathrm{BC}}$ and $F_{c,\mathrm{BC}}$ (nom. par. known), and nominal value $F_c = 2\,\mathrm{N}$.

Figure 3.6: Simulation result of bias compensating parameter estimation with noisy model output, where $e_k \sim \mathcal{N}(0, \sigma_x^2)$ with $\sigma_x = 0.001\,\mathrm{m}$.

The simulation results in Fig. 3.6 show the compelling results of the proposed bias compensation approach. It has to be emphasized that the quality of the bias compensation is independent of its size. The BC-OLS or BC-RLS algorithm delivers consistent or asymptotically unbiased estimation results, even for parameters with a huge least-squares bias, e.g., for the vehicle mass with more than 10% deviation from the nominal value.

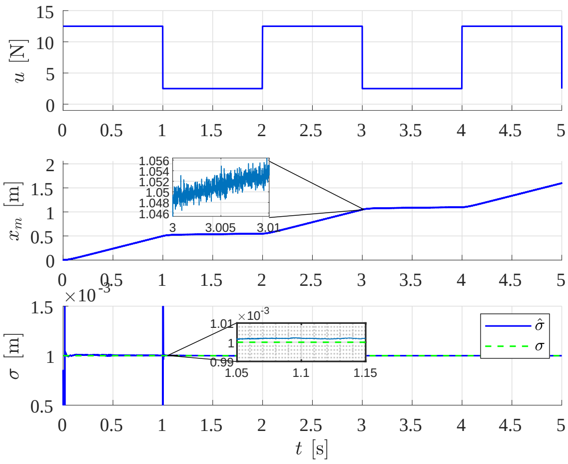
As described in Section 3.4.2, the nominal model parameters, which are necessary for the calculation of the bias, are not known in practice. For this reason, the best available estimates at the respective time are used. In order to find out whether this has a negative influence on the stability of the algorithm, Fig. 3.6 shows, in addition to the estimates obtained by (3.223), e.g., $\hat{m}_{\mathrm{BC}}$, the estimation results one would obtain if the nominal parameter values are known, e.g., $m_{\mathrm{BC}}$. These two signals converge, from which one can conclude that the use of the estimated values does not result in instability issues.

Small filter time constants are necessary for dynamic estimation of model parameters, for example, for real-time monitoring of specific parameters for fault detection. For this reason, the effects of the filter time constant on the estimation results are investigated. Figure 3.7 shows the estimated model parameters using three different filter time constants. The larger the filter time constants, the smaller the cut-off frequency of the low-pass filters. Therefore, the larger the time constant, the more the signal is smoothed, and thus the noise term is suppressed, resulting in more accurate estimation results using OLS. As expected, independent of the chosen filter time constant, the BC-RLS algorithm provides consistent (asymptotically unbiased) results. However, if the time constant is chosen too large, relevant parts of some signals might be lost, or the convergence rate is too slow, e.g., to detect fast parameter changes, making parameter identification imprecise or inapplicable.

#### 3.4.3.1 Comparison with Total Least-Squares Method

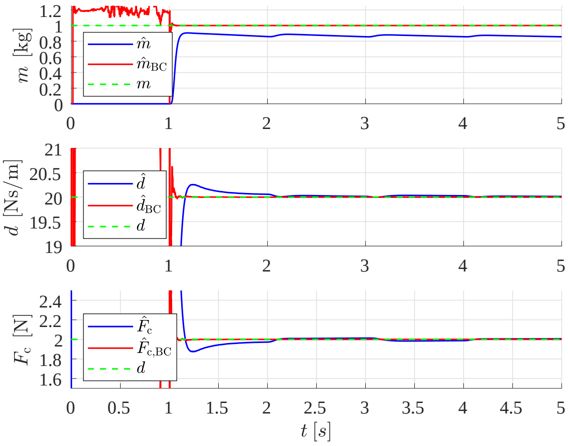
With the ordinary least-squares (OLS) estimator, one obtains unbiased parameter estimates if only the observation vector is noise affected, but the data vector is deterministic. However, in most engineering applications, this assumption is violated, which is why the OLS method generally provides biased results. A well-known linear estimator that consideres random effects in the data vector is the total least-squares (TLS) method. For this reason, a comparison of the proposed bias-compensated (recursive) least-squares algorithm with an iterative total least-squares algorithm, as suggested in [57], is performed.[14] The iterative TLS algorithm is initialized with the OLS solu-

[14] It should be noted that a arbitrary selected TLS algorithm was used here for comparison purposes. Specific TLS algorithms may give more accurate estimation results than the algorithm from [57].

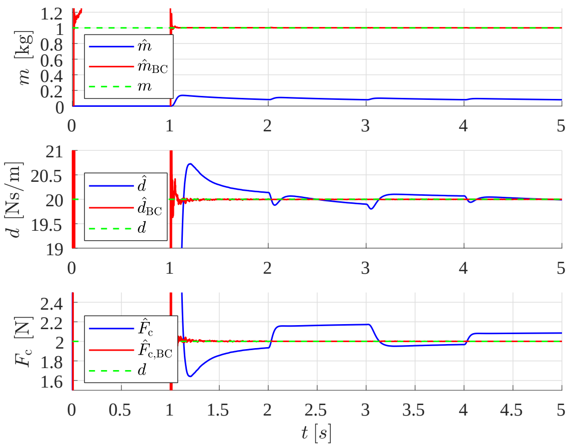



(a) Cart input force $u$, noise affected system output $x_m$ and nominal and (estimated) noise std. deviation $\hat{\sigma}_x$

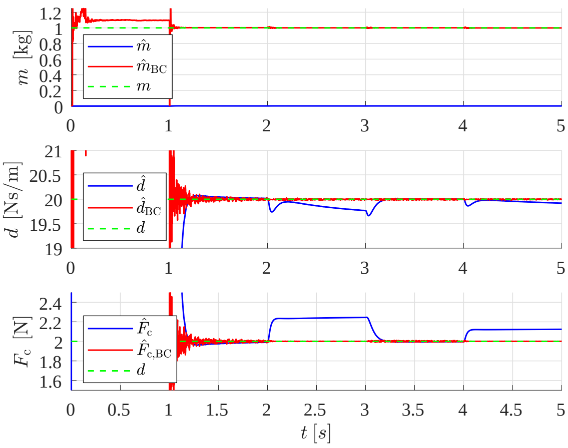
.




(b) Estimated plant parameters with $T_f = 50\,\text{ms}$. Nominal values are dashed green, RLS estimates are blue, and BC-RLS estimates are red.




(c) Estimated plant parameters with $T_f = 15\,\text{ms}$. Nominal values are dashed green, RLS estimates are blue, and BC-RLS estimates are red.




(d) Estimated plant parameters with $T_f = 5\,\text{ms}$. Nominal values are dashed green, RLS estimates are blue, and BC-RLS estimates are red.

Figure 3.7: Comparing RLS and BC-RLS in simulation using different filter time constants $T_f$. The output is noise affected, where $e_k \sim \mathcal{N}\left(0, \sigma_x^2\right)$ with $\sigma_x = 0.001\,\text{m}$.

tion and is iterated until convergence is reached and reads:

- init: $\hat{\mathbf{p}}_{TLS}^0 = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ}$
- i-th iter: $\hat{\mathbf{p}}_{TLS}^i =$

$$\left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \left( \mathbf{W}^\top \mathbf{y}_{LSQ} + \frac{\left(\mathbf{y}_{LSQ} - \mathbf{W}\hat{\mathbf{p}}_{TLS}^{i-1}\right)^\top \left(\mathbf{y}_{LSQ} - \mathbf{W}\hat{\mathbf{p}}_{TLS}^{i-1}\right)}{1 + \left(\hat{\mathbf{p}}_{TLS}^{i-1}\right)^\top \hat{\mathbf{p}}_{TLS}^{i-1}} \right)$$

- end: $||\hat{\mathbf{p}}_{TLS}^i - \hat{\mathbf{p}}_{TLS}^{i-1}|| < \epsilon$

(3.225)

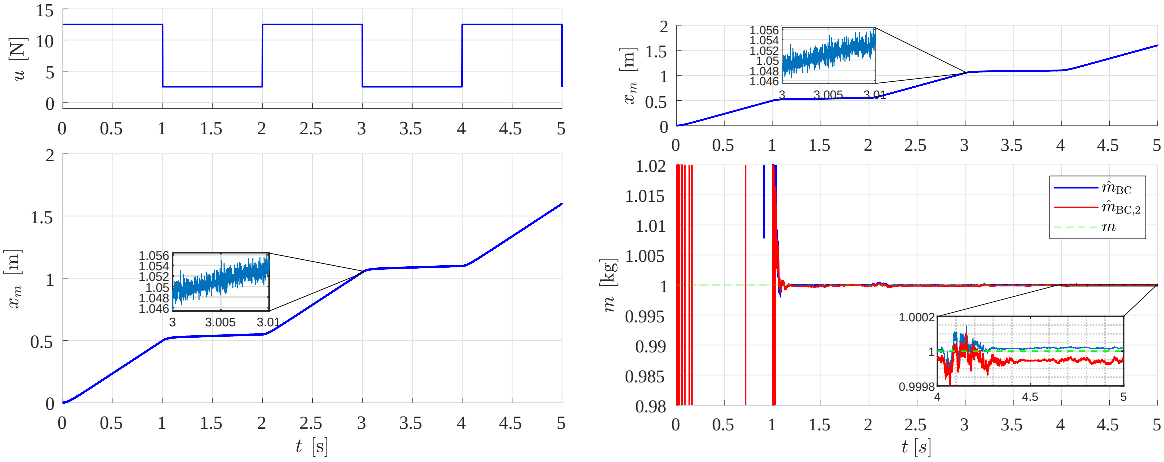
For comparison, the simulation experiment from Fig. 3.6 is used. The TLS significantly improves the estimation result compared to OLS for the parameter $m$, but the estimation result for the friction parameters $d$ and $F_c$ deteriorates (see Tab. 3.1). In summary, the TLS algorithm achieves a less good result than the BC-OLS or BC-RLS.

| nominal parameters | $m = 1\,\mathrm{kg}$ | $d = 20\,\mathrm{Nsm}^{-1}$ | $F_c = 2\,\mathrm{N}$ |
|---|---|---|---|
| estimation method | $\hat{m}$ [kg] | $\hat{d}$ [$\mathrm{Nsm}^{-1}$] | $\hat{F}_c$ [N] |
| OLS | 0.8537 | 20.0055 | 2.0022 |
| BC-OLS | 1.0000 | 19.9995 | 2.0000 |
| TLS (4 iterations) | 0.9588 | 21.7386 | 1.6242 |

Table 3.1: Comparing TLS, OLS and BC-OLS methods assuming noisy model output $x_{m;k} = x_k + e_k$ with additive gaussian white noise, where $e_k \sim \mathcal{N}(\mu, \sigma_x^2)$ with $\mu = 0$ and $\sigma_x = 0.001\,\mathrm{m}$. The sample time is $T = 10\,\mu\mathrm{s}$, and PMF filter time constant is $50\,\mathrm{ms}$.

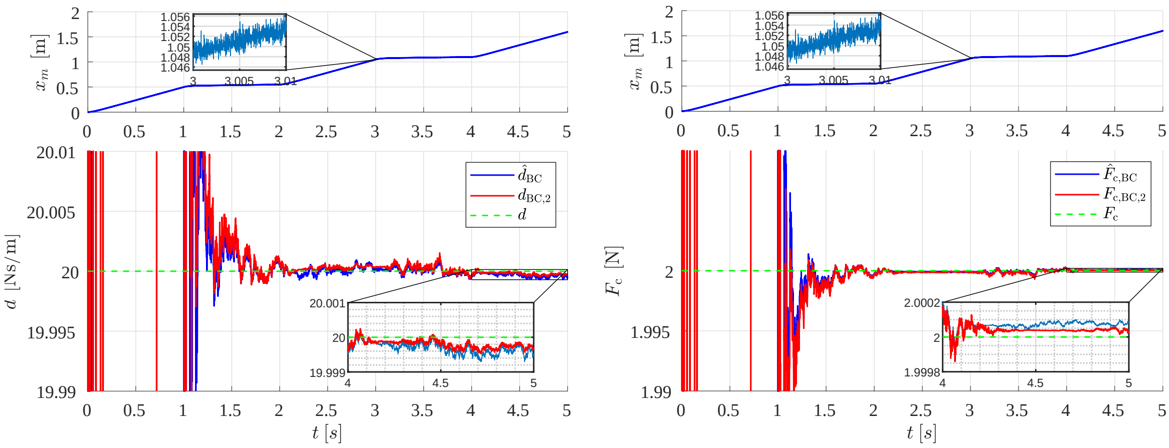
#### 3.4.3.2 Comparison with Other Bias Compensation Methods

For comparisons of the performance of the bias compensation algorithm, discussed in this thesis with other methods, we use the BC-OLS algorithm proposed in [97, 98]. This approach is based on an augmented model by introducing a pre-filter to the model input. To draw conclusions about the estimation bias, one can estimate the parameters of the extended model, where the additional nominal parameter values are known. As the simulation result shows, the results of the two algorithms are quite equivalent.




(a) Cart input force $u$ and noise affected system output $x_m$.

(b) Measured output $x_m$, BC-RLS estimation $\hat{m}_{\mathrm{BC}}$ according to (3.223), BC-RLS estimation $\hat{m}_{\mathrm{BC,2}}$ according to [98], and nominal value $m = 1\,\mathrm{kg}$.




(c) Measured output $x_m$, BC-RLS estimation $\hat{d}_{\mathrm{BC}}$ according to (3.223), BC-RLS estimation $\hat{d}_{\mathrm{BC,2}}$ according to [98], and nominal value $d = 20\,\mathrm{N\,s\,m^{-1}}$.

(d) Measured output $x_m$, BC-RLS estimation $\hat{F}_{\mathrm{c,BC}}$ according to (3.223), BC-RLS estimation $\hat{F}_{\mathrm{c,BC,2}}$ according to [98], and nominal value $F_{\mathrm{c}} = 2\,\mathrm{N}$.

Figure 3.8: Comparison of the BC-RLS algorithm proposed in this thesis and the algorithm introduced in [97, 98] in simulation with noisy model output, where $e_k \sim \mathcal{N}\left(0, \sigma_x^2\right)$ with $\sigma_x = 0.001\,\mathrm{m}$.

## 3.5 Basic Studies on Bias Compensation using Modulating Function Method

As introduced in Section 3.1.2.1.2, to avoid the direct calculation of time derivatives, the derivative operators are shifted from the input signal $u\,(t)$ and output signal $y\,(t)$

to a known modulating function $\varphi(t)$. Considering the fundamental equation

$$\int_0^{T'} \varphi(t)\, y^{(i)}(t)\,\mathrm{d}t = \int_0^{T'} (-1)^i\, \varphi^{(i)}(t)\, y(t)\,\mathrm{d}t \tag{3.226}$$

one obtains an equation suitable for identification in the well-known form

$$y_{LSQ}(t) = \mathbf{w}^\top(t)\,\mathbf{p} \tag{3.227}$$

with

$$\begin{aligned} y_{LSQ}(t) &= \int_{t-T'}^{t} (-1)^n\, \varphi^{(n)}(\tau - t + T')\, y(\tau)\,\mathrm{d}\tau \\ w_{y;i}(t) &= \int_{t-T'}^{t} (-1)^{i+1}\, \varphi^{(i)}(\tau - t + T')\, y(\tau)\,\mathrm{d}\tau\,, \qquad 0 \le i \le n-1 \\ w_{u;j}(t) &= \int_{t-T'}^{t} (-1)^j\, \varphi^{(j)}(\tau - t + T')\, u(\tau)\,\mathrm{d}\tau\,, \qquad 0 \le j \le m_d \end{aligned} \tag{3.228}$$

where

$$\mathbf{w}^\top(t) = \begin{pmatrix} w_{y;0}(t) & \cdots & w_{y;n-1}(t) & w_{u;0}(t) & \cdots & w_{u;m_d}(t) \end{pmatrix} \in \mathbb{R}^{n_p}\,. \tag{3.229}$$

The more or less arbitrary parameter $T'$ denotes the horizon length of the backward parameter estimation. According to Definition 3.1, $\varphi(t)$ is called modulating function (MF) with

$$\varphi^{(i)}(0) = \varphi^{(i)}(T') = 0\,, \quad 0 \le i \le n\,. \tag{3.230}$$

Evaluating (3.227) at a discrete-time instance $t_k = kT$ , $1 \le k \le N$, where $T$ denotes the sample time and $N$ the number of samples, one obtains

$$y_{LSQ;k} = \mathbf{w}_k^\top \mathbf{p} \tag{3.231}$$

with

$$\begin{aligned} y_{LSQ}(t_k) = y_{LSQ;k} &= \int_{t_k-T'}^{t_k} (-1)^n\, \varphi^{(n)}(\tau - t_k + T')\, y(\tau)\,\mathrm{d}\tau \\ w_{y;i}(t_k) = w_{y;i,k} &= \int_{t_k-T'}^{t_k} (-1)^{i+1}\, \varphi^{(i)}(\tau - t_k + T')\, y(\tau)\,\mathrm{d}\tau\,, \quad 0 \le i \le n-1 \\ w_{u;j}(t_k) = w_{u;j,k} &= \int_{t_k-T'}^{t_k} (-1)^j\, \varphi^{(j)}(\tau - t_k + T')\, u(\tau)\,\mathrm{d}\tau\,, \quad 0 \le j \le m_d \end{aligned} \tag{3.232}$$

and constant parameter vector

$$\mathbf{p} = \left(a_0, \cdots, a_{n-1}, b_0, \cdots, b_{m_d}\right)^\top \in \mathbb{R}^{n_p} . \tag{3.233}$$

Evaluating for $N$ different $t_k$, the resulting overdetermined system of equations

$$\underbrace{\begin{pmatrix} y_{LSQ;1} \\ y_{LSQ;2} \\ \vdots \\ y_{LSQ;k} \\ \vdots \\ y_{LSQ;N} \end{pmatrix}}_{=\mathbf{y}_{LSQ}} = \underbrace{\begin{pmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_k^\top \\ \vdots \\ \mathbf{w}_N^\top \end{pmatrix}}_{=\mathbf{W}} \underbrace{\begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \\ b_0 \\ \vdots \\ b_{m_d} \end{pmatrix}}_{=\mathbf{p}} \tag{3.234}$$

where $\mathbf{y}_{LSQ} \in \mathbb{R}^N$, $\mathbf{W} \in \mathbb{R}^{N \times n_p}$, and $\mathbf{p} \in \mathbb{R}^{n_p}$. The optimal parameter estimation in the sense of least-squares reads

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} . \tag{3.235}$$

Applied to the (academic) first-order system with its differential equation

$$\begin{aligned} y(t) &= -T_1 \dot{y}(t) + V u(t) \\ &= \begin{pmatrix} -\dot{y}(t) & u(t) \end{pmatrix} \begin{pmatrix} T_1 \\ V \end{pmatrix} \end{aligned} \tag{3.236}$$

with output $y(t) \in \mathbb{R}$, input $u(t) \in \mathbb{R}$, regressor vector $\mathbf{w}^\top(t) = (-\dot{y}(t), u(t)) \in \mathbb{R}^2$, and parameter vector $\mathbf{p} = (T_1, V)^\top \in \mathbb{R}^2$, one obtains

$$\begin{aligned} &\int_{t-T'}^{t} \varphi(\tau - t + T') \, y(\tau) \, \mathrm{d}\tau = \\ &\qquad \begin{pmatrix} \int_{t-T'}^{t} \dot{\varphi}(\tau - t + T') \, y(\tau) \, \mathrm{d}\tau & \int_{t-T'}^{t} \varphi(\tau - t + T') \, u(\tau) \, \mathrm{d}\tau \end{pmatrix} \begin{pmatrix} T_1 \\ V \end{pmatrix} . \end{aligned} \tag{3.237}$$

By considering noisy measurements

$$y_m(t) = y(t) + e(t) , \quad e(t) \sim i.i.d\left(0, \sigma^2\right) \tag{3.238}$$

and evaluating the continuous functions at discrete time instances $t_k$, one obtains

$$y_{LSQ;k} = \mathbf{w}_k^\top \mathbf{p} + v_k \tag{3.239}$$

with

$$\begin{aligned} y_{LSQ;k} &= \int_{t_k-T'}^{t_k} \varphi\left(\tau - t_k + T'\right) y_m\left(\tau\right) \mathrm{d}\tau \\ \mathbf{w}_k^\top &= \begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(\tau - t_k + T'\right) y_m\left(\tau\right) \mathrm{d}\tau & \int\limits_{t_k-T'}^{t_k} \varphi\left(\tau - t_k + T'\right) u\left(\tau\right) \mathrm{d}\tau \end{pmatrix} \end{aligned} \tag{3.240}$$

and a composite noise term

$$\begin{aligned} v_k &= \int_{t_k-T'}^{t_k} \varphi\left(\tau - t_k + T'\right) e\left(\tau\right) \mathrm{d}\tau - T_1 \int_{t_k-T'}^{t_k} \dot{\varphi}\left(\tau - t_k + T'\right) e\left(\tau\right) \mathrm{d}\tau \\ &= \begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \varphi\left(\tau - t_k + T'\right) e\left(\tau\right) \mathrm{d}\tau & - \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(\tau - t_k + T'\right) e\left(\tau\right) \mathrm{d}\tau \end{pmatrix} \begin{pmatrix} 1 \\ T_1 \end{pmatrix} . \end{aligned} \tag{3.241}$$

Crucial for the calculation of the asymptotic bias according to (3.126) is to determine $\mathrm{E}[\mathbf{w}_k v_k]$. Please note, for better readability, one substitutes $t^\star = \tau - t_k + T'$. Hence, using (3.240) and (3.241), one obtains

$$\begin{aligned} &\mathrm{E}[\mathbf{w}_k v_k] = \\ &\mathrm{E}[\begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) y_m\left(\tau\right) \mathrm{d}\tau \\ \int\limits_{t_k-T'}^{t_k} \varphi\left(t^\star\right) u\left(\tau\right) \mathrm{d}\tau \end{pmatrix} \begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \varphi\left(t^\star\right) e\left(\tau\right) \mathrm{d}\tau & - \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) e\left(\tau\right) \mathrm{d}\tau \end{pmatrix}] \begin{pmatrix} 1 \\ T_1 \end{pmatrix} \\ &= \mathrm{E}[\mathbf{w}_k e_k] \begin{pmatrix} 1 \\ T_1 \end{pmatrix} \end{aligned} \tag{3.242}$$

or

$$\begin{aligned} &\mathrm{E}[\mathbf{w}_k v_k] = \\ &\mathrm{E}[\begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) y_m\left(\tau\right) \mathrm{d}\tau \int\limits_{t_k-T'}^{t_k} \varphi\left(t^\star\right) e\left(\tau\right) \mathrm{d}\tau & - \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) y_m\left(\tau\right) \mathrm{d}\tau \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) e\left(\tau\right) \mathrm{d}\tau \\ \int\limits_{t_k-T'}^{t_k} \varphi\left(t^\star\right) u\left(\tau\right) \mathrm{d}\tau \int\limits_{t_k-T'}^{t_k} \varphi\left(t^\star\right) e\left(\tau\right) \mathrm{d}\tau & - \int\limits_{t_k-T'}^{t_k} \varphi\left(t^\star\right) u\left(\tau\right) \mathrm{d}\tau \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) e\left(\tau\right) \mathrm{d}\tau \end{pmatrix}] \begin{pmatrix} 1 \\ T_1 \end{pmatrix} \\ &= \mathrm{E}[\mathbf{w}_k e_k] \begin{pmatrix} 1 \\ T_1 \end{pmatrix} . \end{aligned} \tag{3.243}$$

Introducing $y_m(t) = y(t) + e(t)$, the first element in the square matrix reads

$$\begin{aligned}\mathrm{E}[\mathbf{w}_k e_k]_{1,1} &= \mathrm{E}[\Big(\int_{t_k-T'}^{t_k} \dot{\varphi}(t^\star)\, y(\tau)\,\mathrm{d}\tau + \int_{t_k-T'}^{t_k} \dot{\varphi}(t^\star)\, e(\tau)\,\mathrm{d}\tau\Big)\Big(\int_{t_k-T'}^{t_k} \varphi(t^\star)\, e(\tau)\,\mathrm{d}\tau\Big)] \\ &= \mathrm{E}[\Big(\int_{t_k-T'}^{t_k} \dot{\varphi}(t^\star)\, y(\tau)\,\mathrm{d}\tau\Big)\Big(\int_{t_k-T'}^{t_k} \varphi(t^\star)\, e(\tau)\,\mathrm{d}\tau\Big)] \\ &\quad + \mathrm{E}[\Big(\int_{t_k-T'}^{t_k} \dot{\varphi}(t^\star)\, e(\tau)\,\mathrm{d}\tau\Big)\Big(\int_{t_k-T'}^{t_k} \varphi(t^\star)\, e(\tau)\,\mathrm{d}\tau\Big)]\,.\end{aligned} \tag{3.244}$$

The time integrals in (3.244) are just ordinary Riemann integrals of a continuous but random function of time $t$. Hence, stochastic calculus rules with a focus on integrals involving random functions are introduced.

**Integral of Stochastic Processes w.r.t Time**

**Proposition 3.1.** The time integral of a continuous random function $X(t) \sim i.i.d\,(0, \sigma^2)$ with respect to $t$ is called Riemann integral and is written as

$$Y = \int_0^{T'} X(t)\,\mathrm{d}t\,. \tag{3.245}$$

The integral approximation

$$Y_N = T \sum_{k=0}^{N-1} X(t_k) \tag{3.246}$$

with $N > 0$ finite summands and sample time $T = \frac{T'}{N}$ , $t_k = kT$ is called (left) Riemann sum. Using two summation variables, for the variance applies

$$\begin{aligned}\mathrm{E}[Y_N^2] &= \mathrm{E}[\Big(T\sum_{k=0}^{N-1} X(t_k)\Big)\Big(T\sum_{j=0}^{N-1} X(t_j)\Big)] \\ &= T^2 \sum_{k=0}^{N-1}\sum_{j=0}^{N-1} \mathrm{E}[X(t_k)\,X(t_j)]\,.\end{aligned} \tag{3.247}$$

Similarly, the covariance of $Y = \int_0^{T'} X(t)\,\mathrm{d}t$ and $V = \int_0^{T'} Z(t)\,\mathrm{d}t$ is

$$\mathrm{E}[Y_N V_N] = T^2 \sum_{k=0}^{N-1}\sum_{j=0}^{N-1} \mathrm{E}[X(t_k)\,Z(t_j)]\,. \tag{3.248}$$

Let $N \to \infty$ or $T \to 0$, the sums converge to a an iterated integral and one obtains the variance

$$\mathrm{E}[Y^2] = \iint_{s,t=0}^{T'} \mathrm{E}[X(t)\,X(s)]\mathrm{d}t\mathrm{d}s \tag{3.249}$$

and covariance

$$\mathrm{E}[YV] = \iint_{s,t=0}^{T'} \mathrm{E}[X(t)\,Z(s)]\mathrm{d}t\mathrm{d}s\,. \tag{3.250}$$

Using (3.250), (3.244) results in

$$\begin{aligned}\mathrm{E}[\mathbf{w}_k e_k]_{1,1} = &\iint_{s,\tau=t_k-T'}^{t_k} \dot{\varphi}(\tau - t_k + T')\,\varphi(s - t_k + T')\,y(\tau)\,\mathrm{E}[e(s)]\mathrm{d}\tau\mathrm{d}s \\ &+ \iint_{s,\tau=t_k-T'}^{t_k} \dot{\varphi}(\tau - t_k + T')\,\varphi(s - t_k + T')\,\mathrm{E}[e(\tau)\,e(s)]\mathrm{d}\tau\mathrm{d}s\end{aligned} \tag{3.251}$$

and simplifies with $\mathrm{E}[e(t)] = 0$ to

$$\mathrm{E}[\mathbf{w}_k e_k]_{1,1} = \iint_{s,\tau=t_k-T'}^{t_k} \dot{\varphi}(\tau - t_k + T')\,\varphi(s - t_k + T')\,\mathrm{E}[e(\tau)\,e(s)]\mathrm{d}\tau\mathrm{d}s\,. \tag{3.252}$$

Approximation of the integral by a finite Riemann sum gives

$$\mathrm{E}[\mathbf{w}_k e_k]_{1,1} \approx T^2 \sum_{i=k-\frac{T'}{T}}^{k} \sum_{j=k-\frac{T'}{T}}^{k} \dot{\varphi}_{i-k+\frac{T'}{T}} \varphi_{j-k+\frac{T'}{T}}\,\mathrm{E}[e_j e_i]\,. \tag{3.253}$$

Considering $e_k \sim i.i.d\,(0, \sigma^2)$, which implies uncorrelated samples

$$\mathrm{E}[e_i e_j] = \begin{cases} \sigma^2\,, & i = j \\ 0\,, & i \neq j \end{cases} \tag{3.254}$$

one obtains

$$\mathrm{E}[\mathbf{w}_k e_k]_{1,1} \approx T^2 \sum_{i=k-\frac{T'}{T}}^{k} \dot{\varphi}_{i-k+\frac{T'}{T}} \varphi_{i-k+\frac{T'}{T}}\,\mathrm{E}[e_i^2] = \sigma^2 T^2 \sum_{i=k-\frac{T'}{T}}^{k} \dot{\varphi}_{i-k+\frac{T'}{T}} \varphi_{i-k+\frac{T'}{T}} \tag{3.255}$$

suitable for implementation purpose. Once again, with $N \to \infty$ or $T \to 0$, the Riemann sum converges to an integral, and one obtains

$$\begin{aligned}\mathrm{E}[\mathbf{w}_k e_k]_{1,1} &= \sigma^2 T \lim_{T\to 0} \sum_{i=k-\frac{T'}{T}}^{k} T \dot{\varphi}_{i-k+\frac{T'}{T}} \varphi_{i-k+\frac{T'}{T}} \\ &= \sigma^2 T \int_{t_k-T'}^{t_k} \dot{\varphi}\left(s - t_k + T'\right) \varphi\left(s - t_k + T'\right) \mathrm{d}s \,.\end{aligned} \tag{3.256}$$

The other elements of $\mathrm{E}[\mathbf{w}_k e_k]$ can be calculated similarly. Finally, one obtains

$$\mathrm{E}[\mathbf{w}_k v_k] \approx \sigma^2 T^2 \begin{pmatrix} \sum_{i=k-\frac{T'}{T}}^{k} \dot{\varphi}_{i-k+\frac{T'}{T}} \varphi_{i-k+\frac{T'}{T}} & -\sum_{i=k-\frac{T'}{T}}^{k} \dot{\varphi}_{i-k+\frac{T'}{T}} \dot{\varphi}_{i-k+\frac{T'}{T}} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ T_1 \end{pmatrix} \tag{3.257}$$

or

$$\mathrm{E}[\mathbf{w}_k v_k] = \sigma^2 T \begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) \varphi\left(t^\star\right) \mathrm{d}s & -\int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) \dot{\varphi}\left(t^\star\right) \mathrm{d}s \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ T_1 \end{pmatrix} \tag{3.258}$$

where $t^\star = s - t_k + T'$. Again, the asymptotic bias is

$$\Delta\mathbf{p} = \mathbf{Q}_W^{-1}\, \mathrm{E}[\mathbf{w}_k v_k] = N \operatorname*{plim}_{N\to\infty} (\mathbf{P})\, \mathrm{E}[\mathbf{w}_k v_k] \tag{3.259}$$

where $\mathbf{P} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1}$.

**Asymptotically Unbiased OLS Estimation for $PT_1$ System using MFM**

**Result 3.6.** The asymptotically unbiased parameter vector reads

$$\begin{aligned}\mathbf{p} &= \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - \Delta\mathbf{p} = \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - N \operatorname*{plim}_{N\to\infty} (\mathbf{P})\, \mathrm{E}[\mathbf{w}_k v_k] \\ &= \operatorname*{plim}_{N\to\infty} \hat{\mathbf{p}} - N \operatorname*{plim}_{N\to\infty} (\mathbf{P})\, \sigma^2 T \begin{pmatrix} \int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) \varphi\left(t^\star\right) \mathrm{d}s & -\int\limits_{t_k-T'}^{t_k} \dot{\varphi}\left(t^\star\right) \dot{\varphi}\left(t^\star\right) \mathrm{d}s \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ T_1 \end{pmatrix}\end{aligned} \tag{3.260}$$

where $\hat{\mathbf{p}}$ and $\mathbf{P}$ are obtained by ordinary least-squares (OLS) estimation

$$\hat{\mathbf{p}} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}_{LSQ} \,. \tag{3.261}$$

In practice, the nominal values of the model parameter $T_1$ and the noise variance $\sigma^2$ in (3.260) are unknown and must be replaced by estimates, c.f., Section 3.4.2. However, in this demonstration example, these values are assumed to be known.

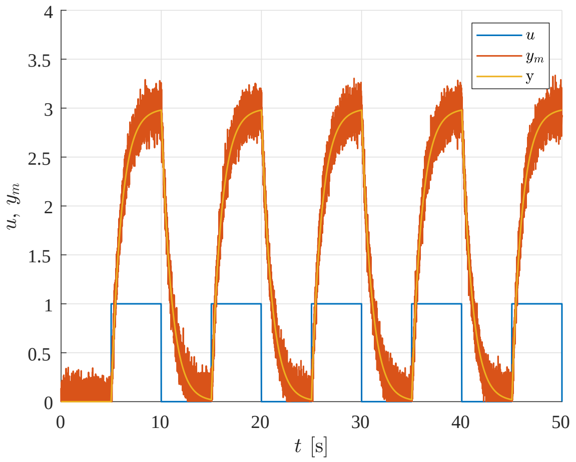
### 3.5.1 Simulation Study

We assume a first-order model (3.236) with nominal parameters $V = 3$ and $T_1 = 1$ and an additive i.i.d. random output noise with distribution $e(t) \sim \mathcal{N}(0, \sigma^2)$, where $\sigma = 0.001$. The output is sampled with sample time $T = 1\,\text{ms}$. As suggested in, e.g., [79], we use a trigonometric modulating function of order $k > 2$

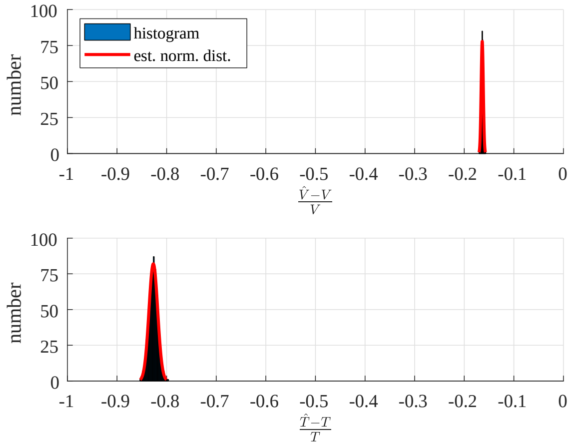
$$\varphi(t) = \left(\sin\left(\frac{k\pi t}{T'}\right)\right)^k \tag{3.262}$$

introduced by [77] with the arbitrary parameters $k = 5$ and $T' = 500T = 0.5\,\text{s}$. Please note that the use of system-specific modulation functions (type, order, etc.) may mitigate the bias problem, i.e., reduces the estimation bias. However, this is explicitly not the subject of this thesis. For example, in [87], an off-line algorithm is proposed to build modulating functions based on the measurements to improve the parameter estimation.

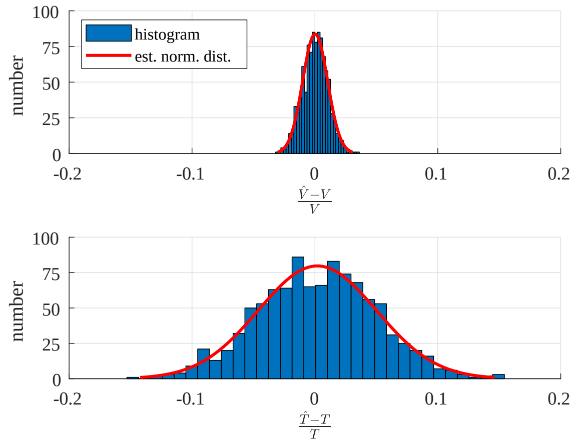
Thousand simulation runs are performed to obtain statistically relevant results, which are summarized in Fig. 3.9. Subfigure 3.9(b) shows the normalized estimation errors using standard least-squares estimation without bias-compensation. The bias caused by the modulating functions is huge, being over 15% for the DC-gain $V$ and even more than 80% for the time constant $T_1$. As expected, the bias compensated estimation according to (3.260) leads to asymptotically unbiased results. Although the simulation runs show that the estimator is asymptotically unbiased, it is not a minimum-variance estimator, i.e., is not efficient according to Definition 3.5.



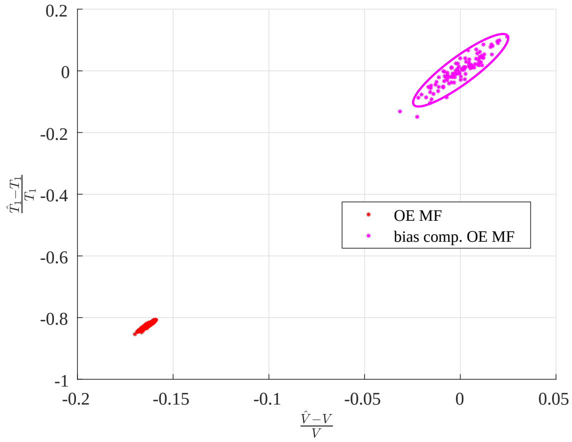

(a) Model input $u$, deterministic output $y$, and noisy output $y_m$.




(b) Normalized estimation errors obtained by MF and OLS.




(c) Normalized estimation errors obtained by MF and BC-OLS.




(d) Parameter covariance (error) ellipse, comparing OLS and BC-OLS. (For sake of clarity, only every tenth value is shown.)

Figure 3.9: The estimation bias, generated by MF, and its calculation and compensation is shown on the simple PT1 model in simulation. Thousand simulation runs are performed to to achieve statistically significant results.

## 3.6 Conclusion

The proposed algorithm to obtain asymptotically unbiased parameter estimates for CT models and the PMF method for noisy outputs was first introduced by [7] for strictly linear models. This thesis works through this approach in detail and is extended to at least slightly nonlinear models, e.g., a 1-D servo positioning system with Coulomb friction. In contrast to other bias-compensating least-squares methods, e.g., [6, 97, 98], the approach presented here does not need any model extensions to get knowledge about the bias, and hence there is no need to estimate additional parameters. The main features of the proposed BC-OLS or BC-RLS algorithm are:

(i) One can obtain an analytical solution of the asymptotic estimation bias, whereby only the PMF filters, their coefficients, and the model parameters, respectively, influence it.

(ii) The algorithm is irrespective of the number of unknown parameters. Only the noise variance must be estimated.

(iii) There are no system extensions necessary. Hence, there is no need to estimate additional parameters, e.g., from the augmented model, getting along with low computational effort, similar to OLS.

(iv) The parameter estimation is robust against noise, i.e., the consistency is mainly independent of the noise model.

(v) An online capable (recursive) algorithm is available.

Furthermore, the same approach is examined to use the more general modulating function method (MFM). Using a simple first-order system as an example shows that the asymptotic bias can be calculated and compensated analogously to that for the PMF approach.

# 4 Applications

## 4.1 Optimal Input Design Applications

### 4.1.1 Engine Model of a Four-Stroke Engine with Port Injection

For the development of control algorithms, highly complex and degree crank angle resolved engine models are unsuitable. For these purposes, so-called mean value models are suitable, which approximate the behavior averaged over an engine cycle and are also suitable for applications in electronic control units (ECUs) due to their lower complexity. One can obtain more detailed explanations about mean value engine models in the literature, e.g., in [125].

A second-order mean value model, consisting of the throttle valve, intake manifold, gas mixing, and the mechanical torque generation, was developed as part of the *RC-LowCAP*[1] project (see Fig. 4.1 for an abstracted overview). Input variables are the

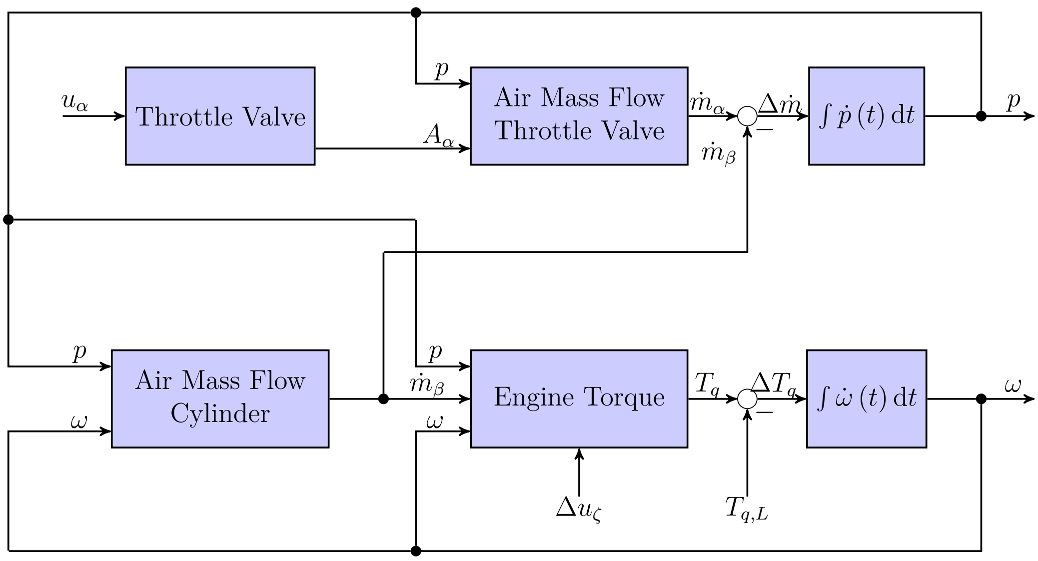



Figure 4.1: Schematic diagram of mean value engine model.

throttle position $u_\alpha \in \mathbb{R}$ and the ignition angle delay $\Delta u_\zeta \in \mathbb{R}$. In contrast to the

[1]FFG-funded COMET project RC-LowCAP.

throttle position, the ignition angle retardation allows a direct intervention on the generated torque and engine speed. State vector $\mathbf{x} = (p, \omega)^\top \in \mathbb{R}^2$ consists of the intake manifold pressure $p$ and the engine speed $\omega$, where the engine speed also represents the model output, i.e., $y = \omega$. The nonlinear model equations read

$$\begin{pmatrix} \dot{p} \\ \dot{\omega} \end{pmatrix} = \begin{pmatrix} -\frac{V_d}{V_m}\frac{\omega}{4\pi}\frac{1}{1+\frac{1}{\lambda\sigma_0}}\left(\frac{V_c+V_d}{V_d} - \frac{V_c}{V_d}\left(\frac{p_{exh}}{p}\right)^{\frac{1}{\kappa}}\left(\gamma_0 + \gamma_1\omega + \gamma_2\omega^2\right)p + \frac{RTp_a}{V_m\sqrt{2RT_a}}\alpha_0\right) \\ \frac{1}{\Theta}\Big(-\left(\beta_0 + \beta_1\omega\right) - \left(p_{exh} - p\right)\frac{V_d}{4\pi} \\ + \frac{RTp_a}{V_m\sqrt{2RT_a}}\alpha_1 u_\alpha \\ + \left(1 - k_\zeta \Delta u_\zeta^2\right)\left(\eta_0 + \eta_1\omega\right)\frac{H_l \dot{m}_\beta}{\omega\sigma_0\lambda} - T_{q,L}\Big) \end{pmatrix} \tag{4.1}$$

where the parameters are explained in Tab. 4.1.[2]

To identify unknown model parameters accurately, the generation of optimal input signals using OID is essential. For example, one can generate optimal input signals for the model inputs throttle position and ignition angle delay to estimate the unknown engine friction coefficients and the thermodynamic efficiency parameters to consider input and state constraints. Hence, the parameter vector for OID is $\mathbf{p}_{OID} = (\eta_0, \eta_1, \beta_0, \beta_1)^\top$. With the vector $\mathbf{p}_0$, containing all nominal parameter values, the Fisher matrix for a large number of samples $N$ according to (2.21) reads

$$\bar{\mathbf{F}}(\mathbf{u}, \mathbf{p}_0) = \frac{1}{N}\sum_{k=1}^{N} \left.\frac{\partial y}{\partial \mathbf{p}_{OID}}\right|_{\mathbf{p}_0, t_k}^\top \left(\sigma^2\right)^{-1} \left.\frac{\partial y}{\partial \mathbf{p}_{OID}}\right|_{\mathbf{p}_0, t_k}, \quad \bar{\mathbf{F}}(\mathbf{u}, \mathbf{p}_0) \in \mathbb{R}^{4\times 4} \tag{4.2}$$

where $\mathbf{u} = (u_\alpha, \Delta u_\zeta)^\top$, $y = \omega$ and $\sigma = 0.001\,\mathrm{rad\,s^{-1}}$. The optimal input signals are

[2] The manifold pressure $p$ is always greater than zero, i.e., the expression $\left(\frac{p_{exh}}{p}\right)^{\frac{1}{\kappa}}$ is not singular.

| | name | unit | nominal value | description |
|---|---|---|---|---|
| plant | $\omega$ | rad/s | - | engine speed |
| | $p$ | Pa | - | intake manifold pressure |
| | $\dot{m}_\beta$ | kg/s | - | air mass flow into cylinder |
| | $u_\alpha$ | % | - | throttle valve opening angle |
| | $\Delta u_\zeta$ | ° | -30 | ignition angle delay,<br>in degree crankshaft angle before TDC |
| | $p_{exh}$ | Pa | - | exhaust pressure |
| | $p_a$ | Pa | - | ambient pressure |
| | $T_{q,L}$ | Nm | - | load torque |
| | $V_d$ | $\mathrm{m}^3$ | - | (effective) engine displacement |
| | $V_c$ | $m^3$ | - | residual cylinder volume TDC |
| | $V_m$ | $\mathrm{m}^3$ | - | volume intake manifold |
| | $R$ | J/ (kgK) | 287 | spefific gas constant (fresh air) |
| | $T$ | K | 317 | air temperature (manifold) |
| | $T_a$ | K | 296 | ambient temperature |
| | $\gamma_0, \gamma_1, \gamma_2$ | - | - | speed dependent volumetric efficiency |
| | $\kappa$ | - | 1.35 | pressure dependent<br>volumetric efficiency |
| | $\alpha_0, \alpha_1$ | - | - | throttle valve coefficients |
| | $\beta_0, \beta_1$ | - | - | engine friction coefficients |
| | $\eta_0, \eta_1$ | - | - | mechanical torque approximation,<br>efficiency coefficients |
| | $\lambda$ | - | - | air-fuel equivalence ratio |
| | $\sigma_0$ | - | 14.7 | stoichiometric air-fuel ratio (gasoline) |
| | $\theta$ | Nm | - | engine moment of inertia |
| | $k_\zeta$ | - | - | ignition angle delay coefficient |
| | $H_l$ | MJ/kg | 45 | lower calorific value,<br>(gasoline) |

Table 4.1: Model parameters for mean value engine model.

obtained by solving

$$
\begin{aligned}
&\max_{u_\alpha, \Delta u_\zeta} \det\left(\bar{\mathbf{F}}\left(\mathbf{u}, \mathbf{p}_0\right)\right) \\
&\begin{pmatrix} \dot{p} \\ \dot{\omega} \end{pmatrix} = \begin{pmatrix} -\frac{V_d}{V_m}\frac{\omega}{4\pi}\frac{1}{1+\frac{1}{\lambda\sigma_0}}\left(\frac{V_c+V_d}{V_d} - \frac{V_c}{V_d}\left(\frac{p_{exh}}{p}\right)^{\frac{1}{\kappa}}\left(\gamma_0+\gamma_1\omega+\gamma_2\omega^2\right)p + \frac{RTp_a}{V_m\sqrt{2RT_a}}\alpha_0\right) \\ \frac{1}{\Theta}\left(-\left(\beta_0+\beta_1\omega\right) - \left(p_{exh}-p\right)\frac{V_d}{4\pi} \right. \\ \qquad + \frac{RTp_a}{V_m\sqrt{2RT_a}}\alpha_1 u_\alpha \\ \left. \qquad + \left(1-k_\zeta\Delta u_\zeta^2\right)\left(\eta_0+\eta_1\omega\right)\frac{H_l\dot{m}_\beta}{\omega\sigma_0\lambda} - T_{q,L}\right) \end{pmatrix} \\
&\mathbf{x}\left(0\right) = \left(109, 2.6\times10^4\right)^\top,\ p\in\left[0, 50\,000\,\mathrm{Pa}\right]\ \omega\in\left[0, 200\,\mathrm{rad\,s^{-1}}\right] \\
&\quad u_\alpha\in\left[0, 100\,\%\right],\ \Delta u_\zeta\in\left[-30\,^\circ\mathrm{ca}, 0\right] \\
&\quad \dot{\mathbf{S}} = \mathbf{J}\mathbf{S}+\mathbf{M},\ t>0,\ \mathbf{S}\left(0\right)=\mathbf{0} \\
&\quad \mathbf{S}_y = \mathbf{S}_{1,\cdot}
\end{aligned}
\tag{4.3}
$$

where $\mathbf{J}\in\mathbb{R}^{2\times2}$ and $\mathbf{M}\in\mathbb{R}^{2\times4}$ are obtained with (2.35) and (2.36), respectively.[3]

Figure 4.2 shows the generated input signals and the corresponding Fisher information.

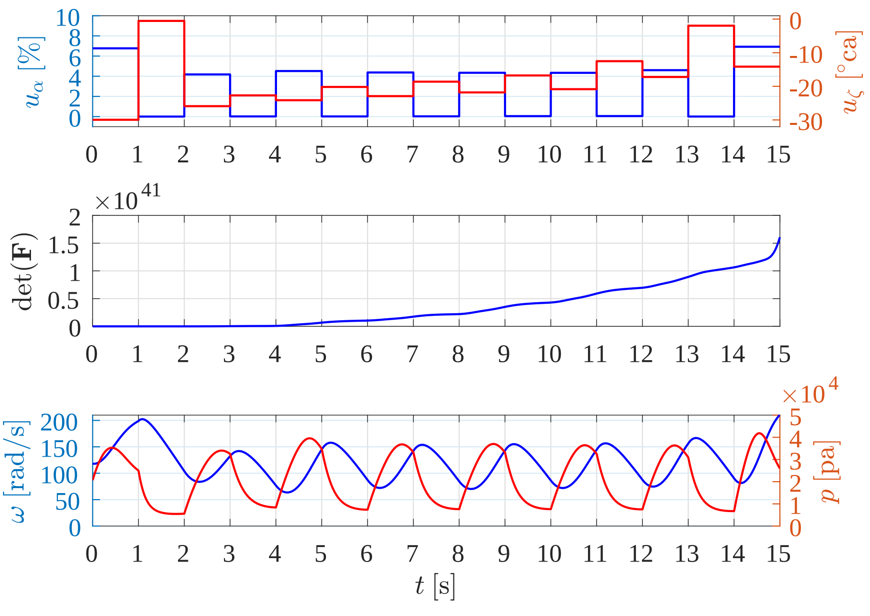



Figure 4.2: Optimal input signals $u_\alpha$ and $\Delta u_\zeta$ for estimating $\eta_0, \eta_1, \beta_0, \beta_1$.

If the engine speed's derivative is determined numerically, one can obtain estimates

[3] Please note: $\mathbf{S}_{i,\cdot}$ denotes the $i$-th row of matrix $\mathbf{S}$.

for the unknown parameters, evaluate

$$\Theta\dot{\omega}_k + \frac{V_d}{4\pi}\left(p_{exh} - p_k\right) = \begin{pmatrix} \frac{H_l}{\lambda\sigma_0}\left(1 - k_\zeta \Delta u_\zeta^2\right)\frac{\dot{m}_\beta}{\omega} & \frac{H_l}{\lambda\sigma_0}\left(1 - k_\zeta \Delta u_\zeta^2\right)\dot{m}_\beta & -1 & -\omega \end{pmatrix} \begin{pmatrix} \eta_0 \\ \eta_1 \\ \beta_0 \\ \beta_1 \end{pmatrix} \tag{4.4}$$

at $N$ discrete-time instances, and use least-squares.

## 4.2 Online Parameter Estimation Applications

### 4.2.1 1-D Rotational Drive with Flywheel

One of the application examples is a 1-D rotational drive, as used in many industrial applications. The system configuration is shown in Fig. 4.3. Performing model re-

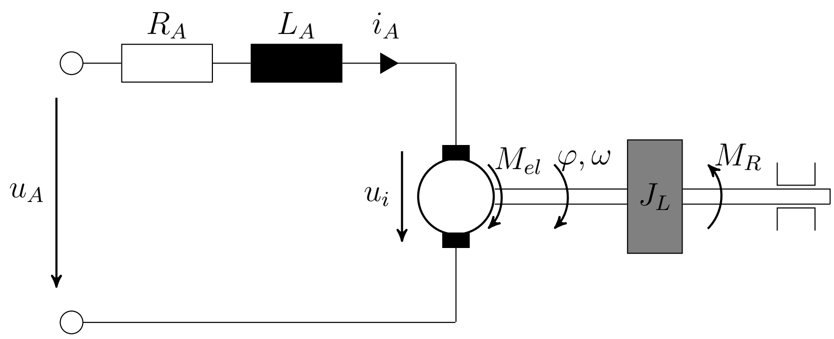



Figure 4.3: Block diagram of 1-D rotational drive with flywheel.

duction to eliminate the dynamics of the electrical subsystem, the respective model equations read

$$\begin{pmatrix} \dot{\varphi} \\ \dot{\omega} \end{pmatrix} = \begin{pmatrix} \omega \\ \frac{\beta}{J}u_A - \frac{\tilde{d}_1}{J}\omega - \frac{M_c}{J}\operatorname{sgn}(\omega) \end{pmatrix} \tag{4.5}$$

with the angular position $\varphi \in \mathbb{R}$, corresponding rotational velocity $\omega \in \mathbb{R}$, input voltage $u_A \in \mathbb{R}$, and friction term $M_R = M_c \operatorname{sgn}(\omega) + d_1\omega$. The known model parameters are the machine constant $k_m$ and the armature resistance $R_A$. Furthermore, for better readability, one can introduce $\beta = \frac{k_m}{R_A}$, $\tilde{d}_1 = d_1 + \frac{km^2}{R_A}$, and $J = J_A + J_L$. The equivalent coefficient of friction consists of two parts: the electrical part and mechanical friction $d_1$. The moment of inertia $J$ consists of the inertia of the electric circuit $J_A$ and the

| | name | unit | nominal value | description |
|---|---|---|---|---|
| plant | $u_A$ | V | - | input voltage |
| | $\varphi$ | rad | - | angular position |
| | $\omega$ | $\frac{\mathrm{rad}}{\mathrm{s}}$ | - | angular velocity |
| | $J_A$ | kgm$^2$ | $41 \times 10^{-7}$ | moment of inertia of electric motor (rotor) |
| | $J_L$ | kgm$^2$ | - | moment of inertia of load (flywheel) |
| | $J$ | kgm$^2$ | - | total moment of inertia, $J = J_A + J_L$ |
| | $d_1$ | $\frac{\mathrm{Nms}}{\mathrm{rad}}$ | - | linear mechanical friction coefficient |
| | $\tilde{d}_1$ | $\frac{\mathrm{Nms}}{\mathrm{rad}}$ | - | equivalent linear rotational friction coefficient |
| | $M_\mathrm{c}$ | Nm | - | rotational Coulomb friction |
| | $k_m$ | $\frac{\mathrm{Nm}}{\mathrm{A}}$ | $38.7 \times 10^{-3}$ | machine constant |
| | $R_A$ | $\Omega$ | 2.079 | armature resistance |
| | $\beta$ | $\frac{\mathrm{Nm}}{\mathrm{V}}$ | $18.6 \times 10^{-3}$ | gain factor voltage to moment, $\beta = \frac{k_m}{R_A}$ |
| PMF | $T$ | s | $1 \times 10^{-3}$ | sample time |
| | $T_f$ | s | $5 \times 10^{-3}$ | PMF filter time constant |
| | $\lambda$ | - | 1 | exp. forgetting factor (BC)-RLS |

Table 4.2: Model parameters and identification algorithm parameters for 1D rotational electric drive.

flywheel $J_L$. The plant parameters

$$\mathbf{p} = \left(J, \tilde{d}_1, M_\mathrm{c}\right)^\top \tag{4.6}$$

are constant but unknown. All parameters and variables are listed in Tab. 4.2.

Applying the PMF approach to eliminate the time derivatives results in

$$\begin{aligned}
\tilde{u}_A(t_k) &= \tilde{u}_{A;k} = \left(g^0_{F,zoh} * u_A\right)_k \\
\hat{\tilde{\varphi}}^{(1)}(t_k) &= \hat{\tilde{\varphi}}^{(1)}_k \approx \left(g^1_{F,tust} * \varphi\right)_k \\
\hat{\tilde{\varphi}}^{(2)}(t_k) &= \hat{\tilde{\varphi}}^{(2)}_k \approx \left(g^2_{F,tust} * \varphi\right)_k \\
\widetilde{\operatorname{sgn}(\omega)}^{(0)}(t_k) &= \widetilde{\operatorname{sgn}(\omega)}^{(0)}_k = \left(g^0_{F,zoh} * \operatorname{sgn}(\omega)\right)_k
\end{aligned} \tag{4.7}$$

where

$$F^i(s) = \frac{s^i}{(1 + sT_f)^2} \quad \text{and} \quad g^i_F(t) = \mathcal{L}^{-1}\left\{F^i(s)\right\}, \quad 0 \le i \le 2 \tag{4.8}$$

and

$$\begin{aligned}
F^i_{zoh}(z) &= \frac{z-1}{z}\mathcal{Z}\left\{\left(\mathcal{L}^{-1}\left\{F^i(s)\frac{1}{s}\right\}\right)\Big|_{t=kT}\right\}, \quad && g^i_{F,zoh}(k) = \mathcal{Z}^{-1}\left\{F^i_{zoh}(z)\right\} \\
F^i_{tust}(z) &= F^i(s') \ \text{ where } s' = \frac{2}{T}\frac{z-1}{z+1}, && g^i_{F,tust}(k) = \mathcal{Z}^{-1}\left\{F^i_{tust}(z)\right\}.
\end{aligned} \tag{4.9}$$

The equation for identification at a single time instance $t_k$ reads

$$\beta \tilde{u}_{A;k} = \begin{pmatrix} \hat{\tilde{\varphi}}_k^{(2)} & \hat{\tilde{\varphi}}_k^{(1)} & \widetilde{\operatorname{sgn}(\omega)}_k^{(0)} \end{pmatrix} \begin{pmatrix} J \\ \tilde{d}_1 \\ M_\mathrm{c} \end{pmatrix} \tag{4.10}$$

where $\operatorname{sgn}(\omega)$ is approximated by computing $\omega$ using discrete quotient and taking sign function.[4]

#### 4.2.1.1 Test Bench Measurements

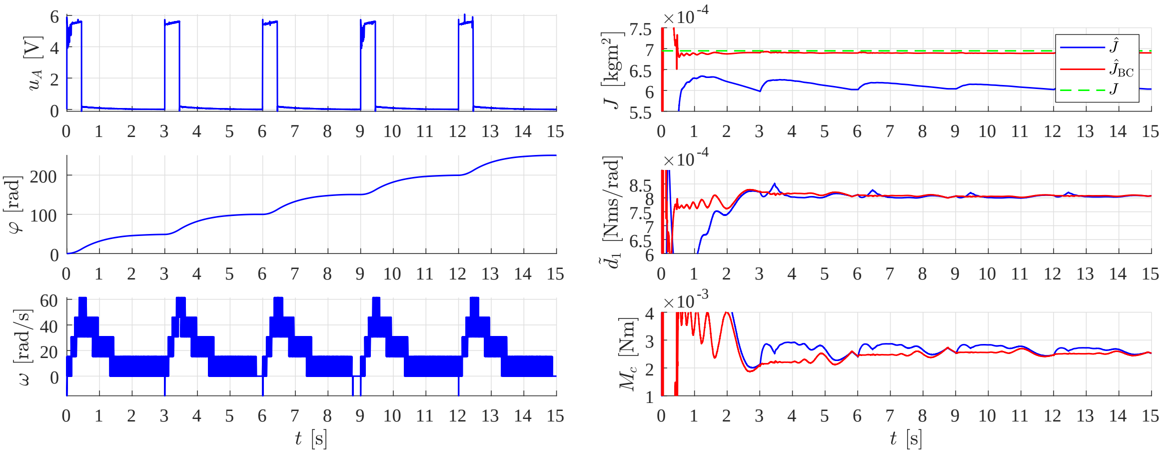
Figure 4.4 shows the parameter estimation results. Smaller filter time constants worsen the standard RLS results, which is understandable because the shorter the time constant of the low-pass filters, the lower the filtering effect on the high-frequency measurement noise. The bias compensated algorithm (BC-RLS) estimates the nominal value of the inertia independent of the filter time constant used.[5] Minimal deviations at larger filter time constants are also visible with the BC-RLS. This phenomenon is because fast signal changes in the input voltage signal caused by non-ideal power electronics are filtered or lost by the larger filter time constants. In summary, one can state that the BC-RLS delivers excellent results for all filter variants and tends to be better even with small filter time constants. In contrast, the standard RLS estimation results are strongly dependent on the selected time constant of the filters. The slight variations in the estimated friction parameters may be caused, among other things, by the violation of the assumptions on the noise conditions. The assumption to approximate encoder quantization noise as additive white noise assumes, among others, random motion. Due to the restriction to one direction of rotation in this measurement, most quantization peaks are positive, which violates the assumptions. Strictly speaking, the approximation of quantization noise as additive white noise is no longer valid.

### 4.2.2 1-D Servo Positioning System

Another practice-oriented example is the 1-D servo positioning system, again used in many industrial applications for any kind of high precision positioning movement, e.g., high-bay storage applications. The system configuration is shown in Fig. 4.5.

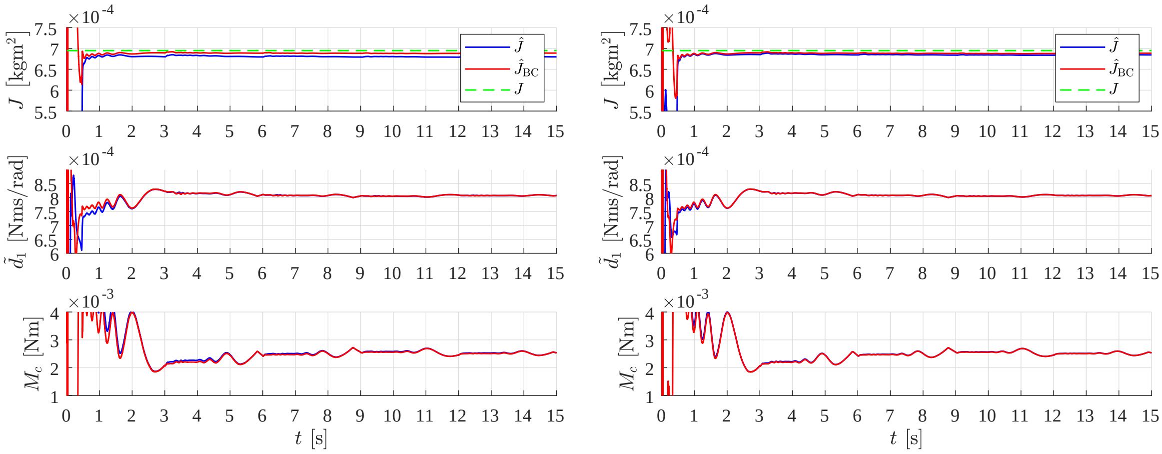
[4] If only positive or negative angular velocities are expected, the term $\operatorname{sgn}(\omega)$ may be replaced by $\pm 1$.

[5] Please note that for the other model parameters, no exact nominal parameters are known.




(a) System input $u_A$, noisy system outputs angular position $\varphi$, and angular velocity $\omega$.

(b) Estimated plant parameters with $T_f = 5\,\mathrm{ms}$. Nominal values are dashed green, RLS estimates are blue, and BC-RLS estimates are red.




(c) Estimated plant parameters with $T_f = 10\,\mathrm{ms}$. Nominal values are dashed green, RLS estimates are blue, and BC-RLS estimates are red.

(d) Estimated plant parameters with $T_f = 15\,\mathrm{ms}$. Nominal values are dashed green, RLS estimates are blue, and BC-RLS estimates are red.

Figure 4.4: Estimated plant parameters, comparing RLS and BC-RLS using different filter time constants $T_f$.

Performing model reduction to eliminate the dynamics of the electrical subsystem, the respective model equations read

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \\ \frac{\beta}{\tilde{m}} u_A - \frac{\tilde{d}_1}{\tilde{m}} v - \frac{F_c}{\tilde{m}} \operatorname{sgn}(v) \end{pmatrix} \tag{4.11}$$

with the translational position $x \in \mathbb{R}$, its corresponding velocity $v \in \mathbb{R}$, input voltage $u_A \in \mathbb{R}$, and friction term $F_R = F_c \operatorname{sgn}(v) + d_1 v$. The known model parameters are the

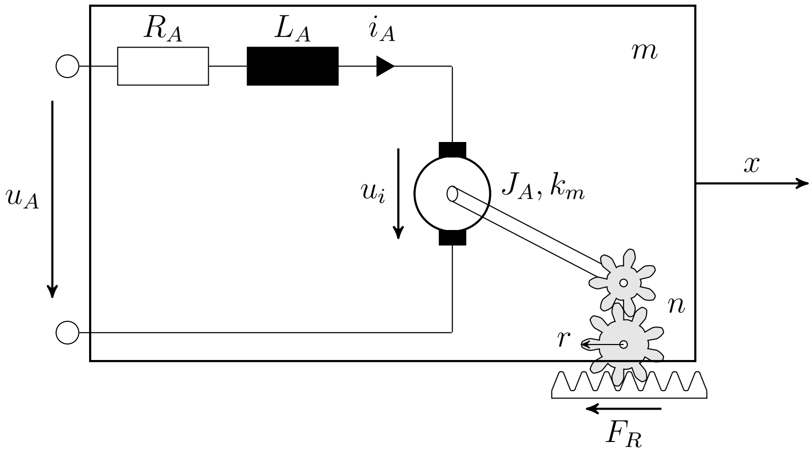



Figure 4.5: Block diagram of 1-D servo positioning system.

machine constant $k_m$, gear ratio $n$, pinion radius $r$, and the armature resistance $R_A$. For better readability, the abbreviations $\beta = \frac{nk_m}{rR_A}$, $\tilde{d}_1 = d_1 + \frac{n^2 km^2}{r^2 R_A}$, and $\tilde{m} = m + J_A \left(\frac{n}{r}\right)^2$ are introduced. The equivalent coefficient of friction consists of two parts: the electrical part and mechanical friction $d_1$. The same thing applies to the equivalent cart mass. Moreover, one assumes that the plant parameters

$$\mathbf{p} = \left(\tilde{m}, \tilde{d}_1, F_{\mathrm{c}}\right)^\top \tag{4.12}$$

are constant but unknown. All parameters and variables are listed in Tab. 4.3.

Applying the PMF approach to eliminate the time derivatives results in

$$\begin{aligned} \tilde{u}_A(t_k) &= \tilde{u}_{A;k} = \left(g^0_{F,zoh} * u_A\right)_k \\ \hat{\tilde{x}}^{(1)}(t_k) &= \hat{\tilde{x}}^{(1)}_k \approx \left(g^1_{F,tust} * x\right)_k \\ \hat{\tilde{x}}^{(2)}(t_k) &= \hat{\tilde{x}}^{(2)}_k \approx \left(g^2_{F,tust} * x\right)_k \\ \widetilde{\operatorname{sgn}(v)}^{(0)}(t_k) &= \widetilde{\operatorname{sgn}(v)}^{(0)}_k = \left(g^0_{F,zoh} * \operatorname{sgn}(v)\right)_k \end{aligned} \tag{4.13}$$

where

$$F^i(s) = \frac{s^i}{\left(1 + sT_f\right)^2} \quad \text{and} \quad g^i_F(t) = \mathcal{L}^{-1}\left\{F^i(s)\right\}, \quad 0 \le i \le 2 \tag{4.14}$$

| | name | unit | nominal value | description |
|---|---|---|---|---|
| plant | $u_A$ | V | - | input voltage |
| | $x$ | m | - | translational position |
| | $v$ | $\frac{\text{m}}{\text{s}}$ | - | translational velocity |
| | $J_A$ | kgm$^2$ | $41 \times 10^{-7}$ | moment of inertia of electric motor (rotor) |
| | $m$ | kg | 0.95 | (mechanical) mass of cart |
| | $\tilde{m}$ | kg | 1.064 | equivalent mass of cart including moment of inertia |
| | $d_1$ | $\frac{\text{Ns}}{\text{m}}$ | 1.94 | linear mechanical friction coefficient |
| | $\tilde{d}_1$ | $\frac{\text{Ns}}{\text{m}}$ | 22 | equivalent linear rotational friction coefficient |
| | $F_c$ | Nm | 0.5 | Coulomb friction |
| | $n$ | - | 1 | gear ratio |
| | $r$ | m | $6 \times 10^{-3}$ | pinion radius |
| | $k_m$ | $\frac{\text{Nm}}{\text{A}}$ | $38.7 \times 10^{-3}$ | machine constant |
| | $R_A$ | $\Omega$ | 2.079 | armature resistance |
| | $\beta$ | $\frac{\text{N}}{\text{V}}$ | 3.1025 | gain factor voltage to force, $\beta = \frac{nk_m}{rR_A}$ |
| PMF | $T$ | s | $1 \times 10^{-3}$ | sample time |
| | $T_f$ | s | - | PMF filter time constant |
| | $\lambda$ | - | 1 | exp. forgetting factor (BC)-RLS |

Table 4.3: Model parameters and identification algorithm parameters for 1-D servo positioning system.

and

$$\begin{aligned} F^i_{zoh}(z) &= \frac{z-1}{z}\mathcal{Z}\left\{\left(\mathcal{L}^{-1}\left\{F^i(s)\frac{1}{s}\right\}\right)\bigg|_{t=kT}\right\}, & g^i_{F,zoh}(k) &= \mathcal{Z}^{-1}\left\{F^i_{zoh}(z)\right\} \\ F^i_{tust}(z) &= F^i(s') \text{ where } s' = \frac{2}{T}\frac{z-1}{z+1}, & g^i_{F,tust}(k) &= \mathcal{Z}^{-1}\left\{F^i_{tust}(z)\right\}. \end{aligned} \tag{4.15}$$

The sample time is $T = 1\,\text{ms}$, and the filter time constant is set to $T_f = 5\,\text{ms}$. The equation for identification at a single time instance $t_k$ reads

$$\beta\tilde{u}_{A;k} = \begin{pmatrix} \hat{\tilde{x}}^{(2)}_k & \hat{\tilde{x}}^{(1)}_k & \widetilde{\operatorname{sgn}(v)}^{(0)}_k \end{pmatrix} \begin{pmatrix} \tilde{m} \\ \tilde{d}_1 \\ F_c \end{pmatrix} \tag{4.16}$$

where $\operatorname{sgn}(v)$ is approximated by computing $v$ using discrete quotient and taking sign function.[6]

[6]If only positive or negative angular velocities are expected, the term $\operatorname{sgn}(v)$ may be replaced by $\pm 1$.

#### 4.2.2.1 PMF Filter Cutoff and Reset

In the practical use of parameter estimation methods, additional tasks arise that have to be solved to improve the parameter estimation result depending on the system or model. These include systems whose behavior is described by switching between different sub-models. A well-known example from the mechanical application area is the servo positioning system with static friction, which has already been used several times in this thesis. Based on this system, both the occurring tasks and a possible solution are described and discussed briefly. Please note that not all model parameters, system states, inputs, and outputs are explained in detail in this section for reasons of clarity. One takes the liberty to refer to Tab. 4.3 in this context. The model equations

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \\ -\frac{\tilde{d}_1}{\tilde{m}} v - \frac{F_c}{\tilde{m}} \operatorname{sgn}(v) + \frac{\beta}{\tilde{m}} u_A \end{pmatrix} \tag{4.17}$$

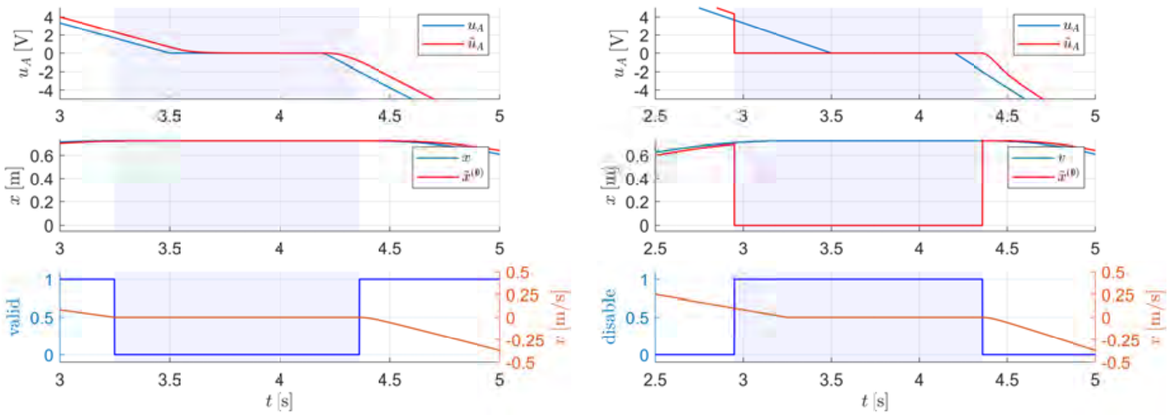
which are transformed into an algebraic equation for identification using suitable methods, are only valid for $v \neq 0$.[7] If the system sticks, it applies

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} . \tag{4.18}$$

Accordingly, from an identification point of view, such invalid standstill phases must not be used for the parameter estimation algorithm. When using the Poisson moment functionals (PMF) approach for eliminating time derivatives, i.e., signal filters, combined with recursive least-squares (RLS) to solve the resulting set of equations, we run into the additional problem of appropriate initialization of the filters restart.

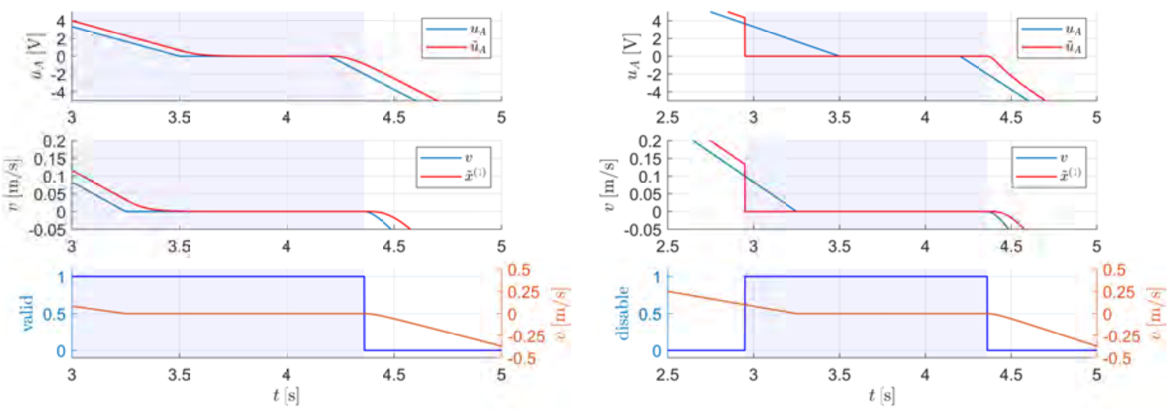
The recursive least squares' inputs must be set to zero when the velocity decays at $v = 0$ or falls below a defined threshold, e.g., $|v| < 0.1\,\mathrm{m\,s^{-1}}$, to cope with measurement uncertainties. Thus, the recursive least-squares algorithm experiences no additional information. Consequently, the estimation result is not falsified. When breaking loose from sticking, one must make sure to reinitialize the PMF filters correctly. Figure 4.6 shows the problem description and an intuitive solution approach.

[7] Please note that in Section 3.4, this practical issue is circumvented by avoiding zero speed phases or minimizing the arising error using PRBS or PRMS input signals with fast zero crossing.




(a) Problem description. Input and velocity signal with the matching PMF filtered signals. During the color-coded standstill phase, i.e. $v = 0$, the equation for identification is not valid.

(b) Solution approach. Disable filter, i.e., set values to zero, when velocity undercuts a threshold value and initialize the filters properly at restart.




(c) Problem description. Input and velocity signal with the matching PMF filtered signals. During the color-coded standstill phase, i.e. $v = 0$, the equation for identification is not valid.

(d) Solution approach. Disable filter, i.e., set values to zero, when velocity undercuts a threshold value and initialize the filters properly at restart.

Figure 4.6: Problem description and solution approach using PMF filters for parameter estimation for switching models, e.g., slip-stick friction model.

As already discussed in detail in Section 3.4, one can use the signal filters (3.175)

$$
\begin{aligned}
F^0(s) &= \frac{1}{(1 + sT_f)^2} \\
F^1(s) &= sF^0(s) \\
F^2(s) &= s^2 F^0(s) = \frac{1}{T_f^2}\left(1 - F^0(s) - 2T_f F^1(s)\right)
\end{aligned}
\tag{4.19}
$$

to eliminate the time derivatives of the measured system signals and finally obtain the algebraic equation for identification[8]

$$\beta \tilde{u}_A^{(0)} = \begin{pmatrix} \tilde{x}^{(2)}(t) & \tilde{x}^{(1)}(t) & \widetilde{\operatorname{sgn}(v)}^{(0)}(t) \end{pmatrix} \begin{pmatrix} \tilde{m} \\ \tilde{d}_1 \\ F_\mathrm{c} \end{pmatrix} \tag{4.20}$$

where

$$\begin{aligned} \tilde{u}_A^{(0)}(t) &= \left(g_F^0 * u_A\right)(t) \;, \qquad g_F^0(t) = \mathcal{L}^{-1}\left\{F^0(s)\right\} \\ \tilde{x}^{(1)}(t) &= \left(g_F^1 * x\right)(t) \;, \qquad g_F^1(t) = \mathcal{L}^{-1}\left\{F^1(s)\right\} \\ \tilde{x}^{(2)}(t) &= \left(g_F^2 * x\right)(t) \;, \qquad g_F^2(t) = \mathcal{L}^{-1}\left\{F^2(s)\right\} \;. \end{aligned} \tag{4.21}$$

Exemplarily, transforming $F^0(s)$ from (4.19) into controllable canonical form obtains the dynamical system

$$\begin{aligned} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ -\frac{1}{T_f^2} & -\frac{2}{T_f} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u \\ y &= \begin{pmatrix} \frac{1}{T_f^2} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned} \tag{4.22}$$

with filter state variables $x_1$ and $x_2$. The other filters are transformed analogously and differ only in the system output. The relation

$$\tilde{u}^{(0)}(t) = \left(g_F^0 * u\right)(t) \;, \quad g_F^0(t) = \mathcal{L}^{-1}\left\{F^0(s)\right\} \tag{4.23}$$

allows the interpretation of the system states

$$\begin{aligned} x_1(t) &= T_f^2 \tilde{u}^{(0)}(t) \\ x_2(t) &= T_f^2 \frac{\mathrm{d}}{\mathrm{d}t}\left\{\tilde{u}^{(0)}(t)\right\} \;. \end{aligned} \tag{4.24}$$

Assuming a sufficiently long standstill before the filter restart[9], for steady-state applies

[8] For application to discrete-time measurement data or for implementation purposes, the filters must be discretized appropriately. For elaborating on the filter reset problem, discretization is not necessary, which is why one remains here in the continuous-time domain. For details on discretization, the kindly reader is referred to Section 3.4.

[9] For slow creep or jumping through zero speed or restart filters at a predefined speed (deadzone around zero velocity), the filters' start values must be calculated differently.

$$x_1(t) = T_f^2 \tilde{u}^{(0)}(t) = T_f^2 u(t)$$
$$x_2(t) = T_f^2 \frac{\mathrm{d}}{\mathrm{d}t}\left\{\tilde{u}^{(0)}(t)\right\} = T_f^2 \frac{\mathrm{d}}{\mathrm{d}t}\{u(t)\} \ . \tag{4.25}$$

Therefore, the filter dynamic system's state variables are directly proportional to the filter's input variable and its first derivative. By filtering servo position $x$, the filter state variables are directly proportional to the position and directly proportional to the velocity, i.e.,

$$x_1(0) = T_f^2 x$$
$$x_2(0) = 0 \ . \tag{4.26}$$

Looking at the equation for identification evaluated at $t = 0$ (filter restart)

$$\beta \tilde{u}_A^{(0)}(0) = \tilde{m}\tilde{x}^{(2)}(0) + \tilde{d}_1 \tilde{x}^{(1)}(0) + F_{\mathrm{c}} \widetilde{\mathrm{sgn}(v)}^{(0)}(0) \tag{4.27}$$

and taking into account $\tilde{x}^{(1)}(0) = \tilde{x}^{(2)}(0) = 0$, one obtains

$$\beta \tilde{u}_A^{(0)}(0) = F_{\mathrm{c}} \widetilde{\mathrm{sgn}(v)}^{(0)}(0) \ . \tag{4.28}$$

Hence, to fulfill (4.28) and to be independent of the unknown parameter $F_{\mathrm{c}}$, one may set the initial conditions $x_1(0)$ and $x_2(0)$ of both, the dynamical systems filtering the input voltage and sign function at the restart to

$$x_1(0) = 0$$
$$x_2(0) = 0 \ . \tag{4.29}$$

One can show the differences in parameter estimation results with a first simulation experiment neglecting measurement noise or encoder quantization effects. Of course, for implementation purposes, the derived dynamical systems or filters are discretized appropriately, and the corresponding initial states must be converted. Again, the kindly reader is referred to Section 3.4. We compare the standard approach, i.e., the filters are active throughout, with the solution presented here, i.e., the filters are disabled when the speed falls below a minimum value $|v| < 0.1\,\mathrm{m\,s^{-1}}$ and are restarted and reinitialized when the velocity changes starting from zero, i.e., $v \neq 0$. Standard pulse or ramp signals serve as input signals (compare Section 2.2), the sampling time is $T = 1\,\mathrm{ms}$, the filter time constant is $T_f = 50\,\mathrm{ms}$, and the nominal model parameters for basic simulation experiments are $\tilde{m} = 1\,\mathrm{kg}$, $\tilde{d}_1 = 20\,\mathrm{N\,s\,m^{-1}}$, $F_{\mathrm{c}} = 2\,\mathrm{N}$.

Figure 4.7(a) shows a pulse-shaped input voltage signal with the respective position and velocity. In the third subplot, one can see where the signal filters or the recursive least squares algorithm should be disabled. In Fig. 4.7(b), one can see the corresponding parameter estimation values. For pulse-shaped input signals, the differences are small. Slight deviations from the nominal values can be explained by the approximative character of the discretized filters or dynamic systems. It is striking that the estimated parameter values do not converge to a final value without eliminating these weak areas. Similar applies when using ramped voltage waveforms, c.f. Fig. 4.7(c) and Fig. 4.7(d). The standard approach deviations without filter reset are even significantly more extensive than for pulse-shaped voltage waveforms. The reason is that the invalid signal components have a more substantial effect in the colored areas due to the less steep edges. Additionally, compared to pulse-shaped signals, ramp-shaped excite fewer frequencies or a smaller frequency spectrum due to the lower slope.

The next step is to add an output noise to the simulation, which means $x_m(t) = x(t) + e(t)$ with $e \sim \mathcal{N}(0, \sigma_x)$, where $\sigma_x = 0.001\,\mathrm{m}$. Figure 4.8 and Fig. 4.9 show the effects of measurement noise on both the standard PMF approach and the filter reset approach, each with standard RLS and BC-RLS for pulse-shaped and ramp-shaped input signals. As expected, the measurement noise significantly worsens the parameter estimation results using standard RLS. Differences between the two PMF approaches and standard RLS are not evident (c.f., Fig. 4.8(b) or Fig. 4.9(b)). The situation is different when using the BC-RLS. Figure 4.9(c) shows that the BC-RLS performs even worse due to the invalid areas than the corresponding RLS for ramp-shaped signals, while the estimation results are quite good for pulsed signals with and without filter reset, see Fig. 4.8(c). The combination of filter reset and bias compensation leads to promising results for puls-shaped and ramp-shaped inputs. The estimated values converge asymptotically towards the nominal values, c.f. Fig. 4.8(c) or Fig. 4.9(c).

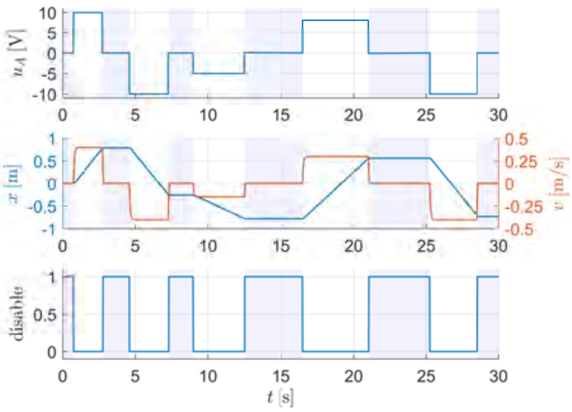
Open issues which could be addressed based on these results are:

(i) appropriate filter initialization if there occur no standstill phases, i.e., zero-crossing

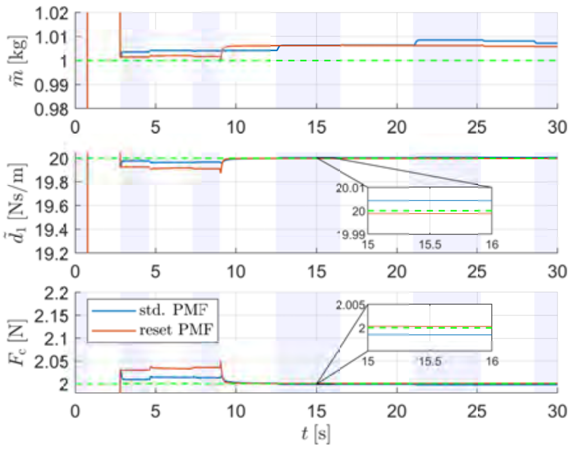
(ii) appropriate filter initialization if the restart should be at a certain velocity unequal to zero, e.g. $|v| > 0.1\,\mathrm{m\,s^{-1}}$ instead of $|v| > 0$ used here

#### 4.2.2.2 Test Bench Measurements

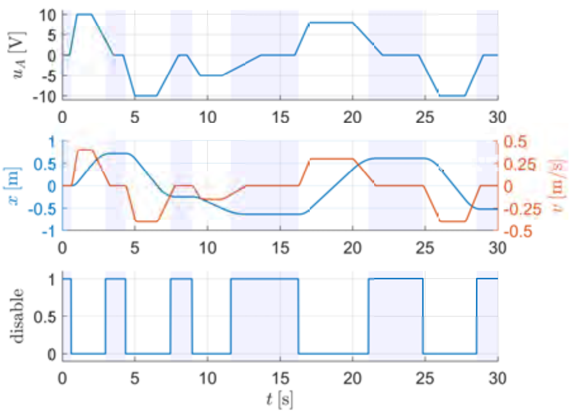
The combination of PMF filter reset and BC-RLS, which was successfully tested in the simulation, is now being evaluated on test bench measurement data from a laboratory test bench, see Fig. 4.10. Besides, measurement data using an OID input signal



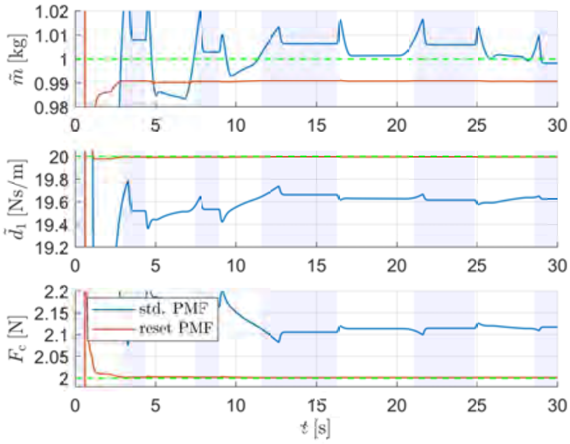

(a) Arbitrarily pulse-shaped system input $u_A$, system outputs $x$, $v$, and filter flag. The filter flag is set when the velocity falls below $0.1\,\mathrm{m\,s^{-1}}$ during deceleration, i.e., $|v| < 0.1\,\mathrm{m\,s^{-1}}$.




(b) Estimated parameter values using pulse-shaped input. Nominal values are dashed green, RLS estimates with standard PMF filters are blue, and RLS estimates with enhanced, i.e., reset, PMF filters are red.



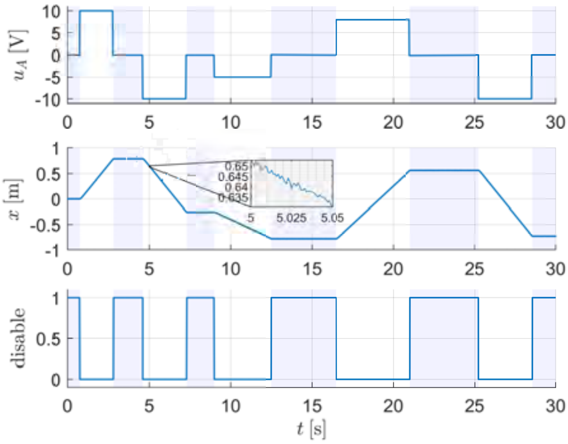

(c) Arbitrarily ramp-shaped system input $u_A$, system outputs $x$, $v$, and filter flag. The filter flag is set when the velocity falls below $0.1\,\mathrm{m\,s^{-1}}$ during deceleration, i.e., $|v| < 0.1\,\mathrm{m\,s^{-1}}$.



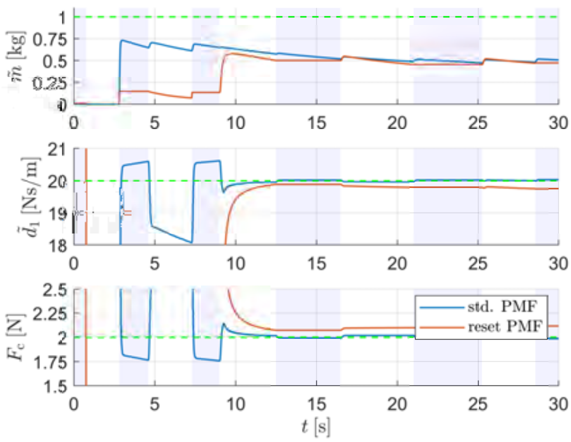

(d) Estimated parameter values using ramp-shaped input. Nominal values are dashed green, RLS estimates with standard PMF filters are blue, and RLS estimates with enhanced, i.e., reset, PMF filters are red.

Figure 4.7: Pulse-shaped or ramp-shaped input signals and corresponding estimated plant parameters with filter time constant $T_f = 50\,\mathrm{ms}$ and sample time $T = 1\,\mathrm{ms}$. In the areas colored light blue, the identification equation is invalid and the PMF filters are turned off.

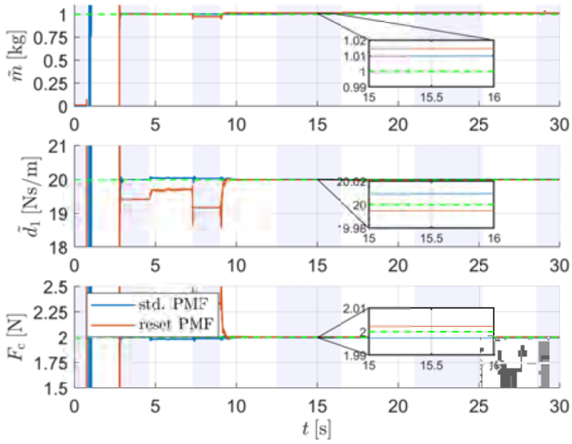
similar to Section 2.6.2 is generated and evaluated. To obtain the nominal electrical parameters, i.e., armature resistance and machine constant, one operates the dc-drive in generator mode. To identify the servo's mechanical friction, one can pull the drive, measuring the tractive force. The nominal parameters determined in this way can be




(a) Pulse-shaped system input $u_A$, noisy system output $x \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.001\,\mathrm{m}$, and filter flag. The filter flag is set when $|v| < 0.1\,\mathrm{m\,s^{-1}}$ during deceleration.




(b) Estimated parameter values. Nominal values are dashed green, RLS estimates with standard PMF are blue, and RLS with PMF reset are red.

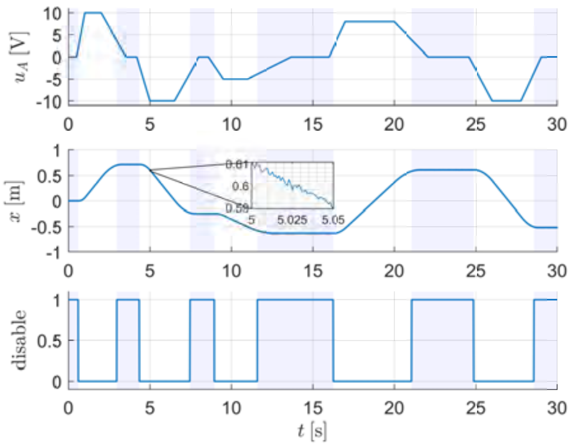



(c) Estimated parameter values. Nominal values are dashed green, BC-RLS estimates with standard PMF are blue, and BC-RLS with PMF reset are red.

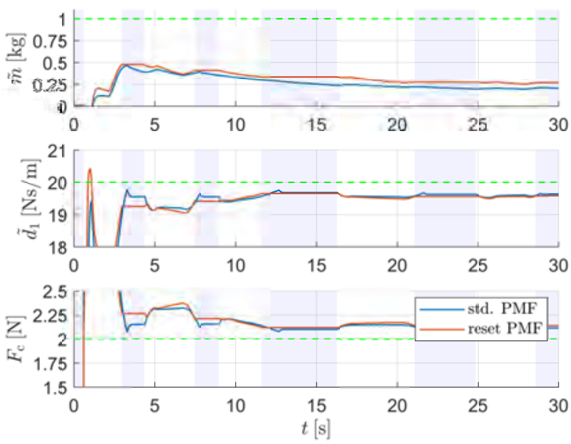
Figure 4.8: Pulse-shaped input signals and estimated plant parameters (RLS vs. BC-RLS) with filter time constant $T_f = 50\,\mathrm{ms}$ and sample time $T = 1\,\mathrm{ms}$. In the areas colored light blue, the identification equation is invalid.

found in Tab. 4.3, e.g., $\tilde{m} = 1.064\,\mathrm{kg}$, $\tilde{d}_1 = 22\,\mathrm{N\,s\,m^{-1}}$, $F_c = 0.5\,\mathrm{N}$.

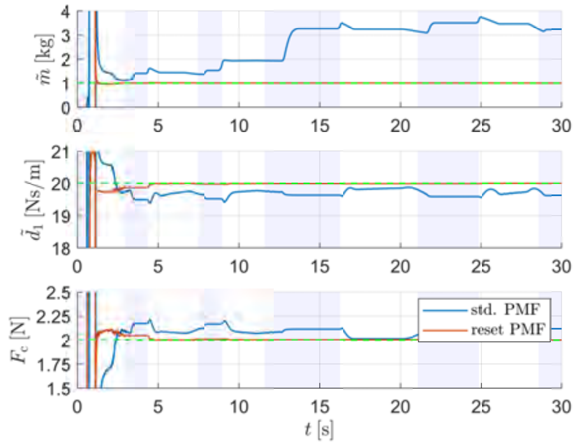
Again, one sets the PMF filter time constant to $T_f = 50\,\mathrm{ms}$ and the sample time $T = 1\,\mathrm{ms}$. Figure 4.11 shows the measurement results using pulse-shaped input signals. The algorithm estimates the unknown parameters quite well. The variations from each other and the nominal parameter values are mainly due to model uncertainties, e.g., the mechanical design of the test bench. For example, mechanical friction is not




(a) Ramp-shaped input $u_A$, noisy system output $x \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.001\,\mathrm{m}$, and filter flag. The filter flag is set when $|v| < 0.1\,\mathrm{m\,s^{-1}}$ during deceleration.




(b) Estimated parameter values. Nominal values are dashed green, RLS estimates with standard PMF are blue, and RLS with PMF reset are red.

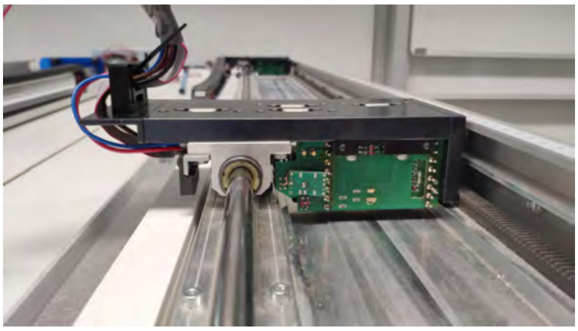



(c) Estimated parameter values. Nominal values are dashed green, BC-RLS estimates with standard PMF are blue, and BC-RLS with PMF reset are red.

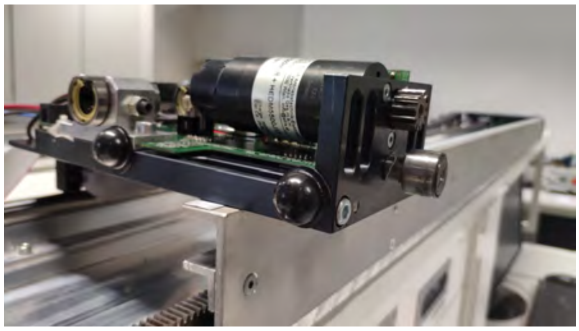
Figure 4.9: Ramp shaped input signals and estimated plant parameters (RLS vs. BC-RLS) with filter time constant $T_f = 50\,\mathrm{ms}$ and sample time $T = 1\,\mathrm{ms}$. In the areas colored light blue, the identification equation is invalid.

constant over the entire range of motion and seems direction-dependent. Additionally, the non-ideal gearbox distorts the results.

Similarly, when using the generated optimal excitation signal. Figure 4.12 shows that the parameter $d_1$ converges better, while the mass is quite similar to Figure 4.11. One may expect that the estimation result would even become a bit more precise by extending the filter reset approach to signals without zero velocity phases. An extension of this concept to signals without standstill phases, e.g., suitable for PRMS-

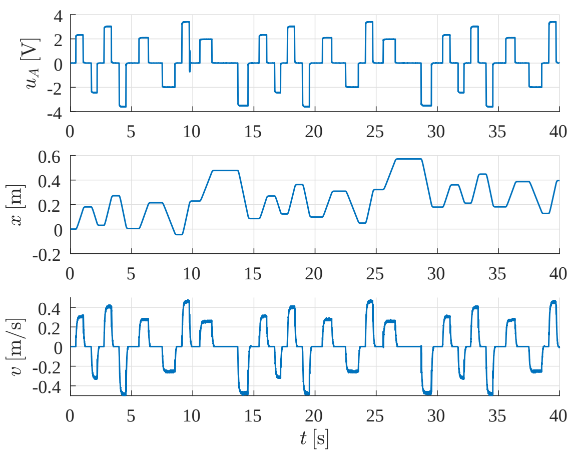


(a) Servo with rack and rail system.

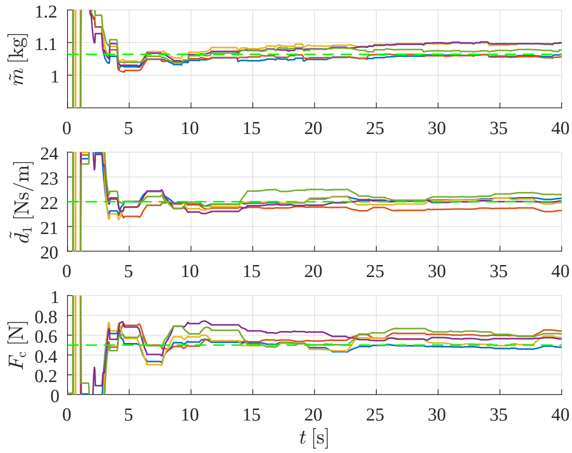


(b) Drive, encoder, ball bearings, gearbox and pinion.

Figure 4.10: Test bench of 1-D servo positioning system.




(a) System input $u_A$, measured outputs position $x$, and velocity $v$. For illustration purpose, one measurement is shown as an example.

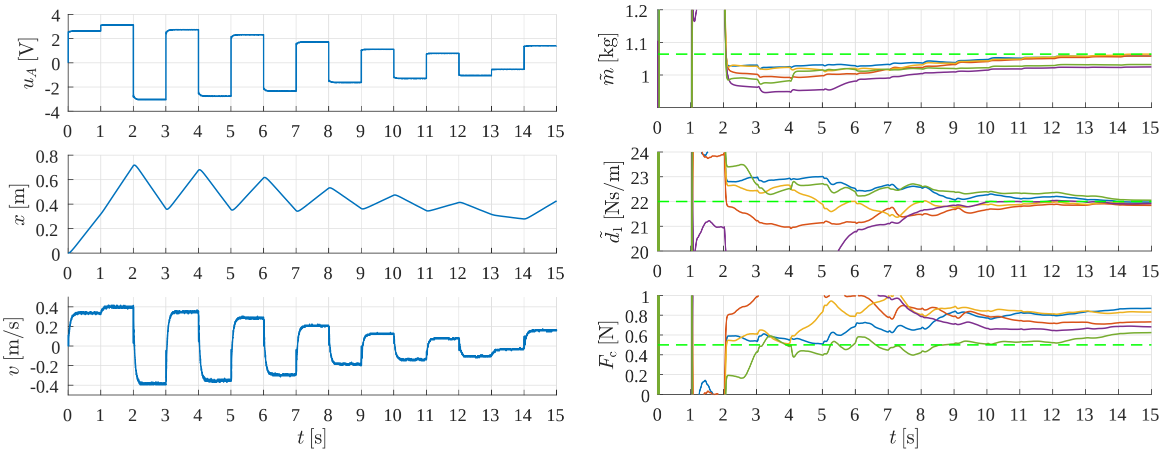



(b) BC-RLS estimated plant parameters with $T_f = 50\,\text{ms}$ for five measurement runs. Nominal values are dashed green.

Figure 4.11: Five test bench measurements. Estimated parameter values are obtained using BC-RLS algorithm for pulse-shaped input voltage with PMF filter reset.

like OID signals, is the task of further work.

Analogous to Fig. 4.9, we also use ramp-shaped voltage signals on the test bench experiments. Figure 4.13 shows the results of five ramps of different steepness with decreasing rising times from 200 ms to 0 ms (pulse-shaped). Due to inadequacies in the mechanics of the test bench, in contrast to the simulation, the estimation results sometimes deviate strongly from the nominal parameters. One can conclude that the less steep the slopes, the more significant the deviations of the estimated parameters from the nominal values, which is especially true for the servo mass. The combination

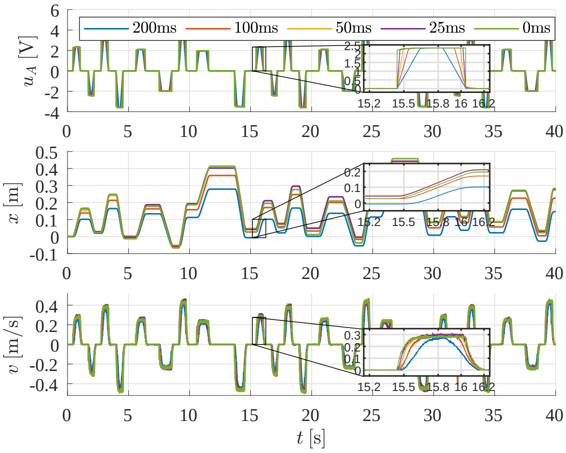



(a) System input $u_A$, measured outputs position $x$, and velocity $v$.

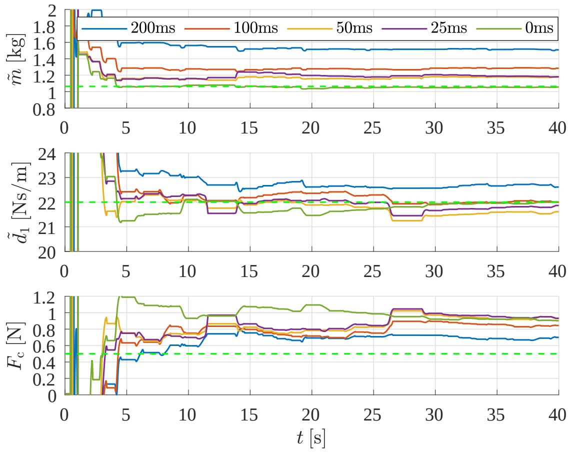
(b) Estimated plant parameters with $T_f = 50\,\mathrm{ms}$. Nominal values are dashed green, BC-RLS estimates are blue.

Figure 4.12: Five test bench measurements. Estimated plant parameters using BC-RLS algorithm for OID-generated input voltage.

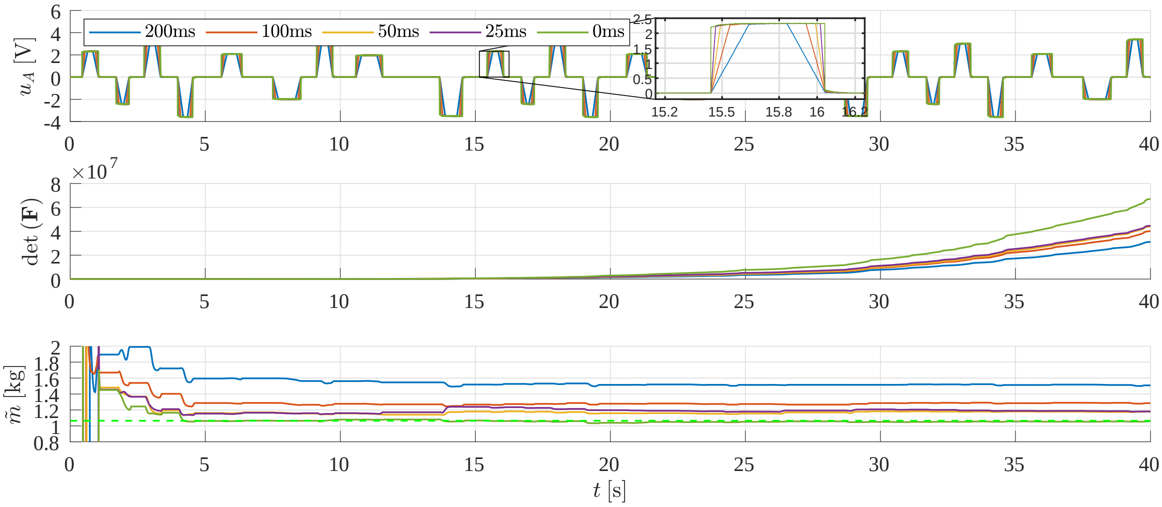
of lower acceleration force and model inaccuracies or model errors leads to these estimation errors. Figure 4.13(c) shows the correlation between the quality of parameter estimation or quality of the excitation signal and the Fisher information. Again, the higher the information content of the input signal, the better the estimation result. This insight again closes the circle and confirms the meaningfulness of the optimal input design (OID) in Chapter 2.




(a) System input $u_A$, measured outputs position $x$, and velocity $v$ for input signals with different rise times.




(b) Estimated plant parameters with different input signal rise times and $T_f = 50\,\text{ms}$. Nominal values are dashed green, BC-RLS estimates are blue.




(c) System input $u_A$, Fisher information det $(\mathbf{F})$, and estimated cart mass $\hat{m}_{BC}$ for different rise times.

Figure 4.13: Five test bench measurements. Estimated plant parameters using BC-RLS algorithm for ramp-shaped input voltage with different rise times.

# 5 Conclusion & Outlook

This chapter summarizes the most challenging tasks, solution approaches, and results of this work and some possible extensions based on them. Short summaries can also be found at the end of each chapter. For reasons of better clarity, these are joined here.

In Chapter 2, it is shown that DoE methods can improve the result of subsequent parameter estimation approaches, independent of the method used. The approach investigated is based on the FIM and thus on local parameter sensitivities. A simple first-order system is used as an introductory striking example, whose model parameters can be calculated from two measuring points of a noisy step response analytically. The task is to choose the measurement points optimally in parameter estimation, i.e., to choose the measurement points so that the model parameters can be determined as accurately as possible. For this simple example, the approach for determining the optimal measurement time points based on the Fisher Information matrix and the subsequent estimation's improvement is graphically illustrated. Often DoE is used to generate informative input signals tailored to the particular experiment. In these cases, one may speak of OID. Using the practical example of a high-speed positioning system, the generation of optimal input signals for accurate parameter estimation is carried out while complying with existing restrictions on the input, output, and state variables for both the linear and nonlinear model with static friction. Simulation results show that such optimal excitation signals significantly improve parameter estimation quality, and specific knowledge about the experiment can be automatically incorporated into the generation of excitation signals. In case of substantial parameter uncertainties, one recommend an iterative process of OID and parameter estimation. In a nutshell: by choosing optimal excitation signals, the accuracy of the subsequent parameter estimation can be increased without taking appropriate methods on the estimation side, e.g., algorithms for bias compensation. The studied method is based on local sensitivities. Mainly when applied to nonlinear systems, this may not be suitable. The possibility of extending the methodology to global sensitivities is briefly indicated in this thesis but not further explored. Hence, it is a possible useful extension of this work.

The second central part of this thesis deals with parameter estimation of continuous-time dynamical systems. The focus is on online algorithms, which is why the combination of PMF and Recursive Least Squares is treated in this work. The application of methods to avoid the direct computation of time derivatives of measurement signals, e.g., PMF, cause a so-called bias in parameter estimation with least-squares in case of noisy signals. The main focus is the theoretical investigation of existing approaches to estimate this bias for strictly linear systems and the extension to slightly nonlinear systems. The advantages of the algorithm studied are:

(i) One can obtain an analytical solution of the asymptotic estimation bias, whereby only the PMF filters and their coefficients, respectively, influence it.

(ii) The Algorithm is irrespective of the number of unknown parameters. Only the noise variance must be estimated.

(iii) Low computing effort, similar to standard OLS, because no system extensions are necessary. Hence, there is no need to estimate additional parameters, e.g., from the augmented model.

(iv) The parameter estimation is robust against noise, i.e., the consistency is independent of the noise model.

(v) An online capable (recursive) algorithm is available.

The same approach is also extended from PMFs to the more general class of MFs (to eliminate derivatives). First simulation tests show that one can obtain asymptotically unbiased parameter estimates.

In Chapter 4, the investigated methods are applied to practical applications. For example, an optimal throttle position and ignition angle delay signal are calculated to identify a mean value engine model's parameters. Investigations on unbiased parameter estimation are carried out on two mechatronic systems, namely the DC-drive with a rotational flywheel and the multiply used 1D-positioning servo using test-bench measurements. A practical problem of parameter estimation is investigated in Section 4.2.2.1 using the nonlinear positioning system as an example. If the system leaves the horizon covered by the identification equation, the estimation algorithm must discard this data. When the system re-enters the valid area, one must ensure that the PMF filters are appropriately initialized. For the 1D-servo with static friction, the identification equation is only valid for velocity not equal to zero. In sticking, the equation is invalid, and the PMF filters must be disabled and re-initialized properly when breaking loose from sticking. Simulation studies show that the estimated values deviate significantly from the nominal parameters without appropriate action, especially when using the bias compensation algorithm. In this work, a particular case of filter re-initialization, which requires a sufficiently long standstill phase before breaking free from sticking, is investigated and validated in simulation. A more general approach, which also considers slow creep or jumping through zero velocity, would be an essential and useful extension to this work.

# A Appendix

Section 3.3 discusses, among other things, the statistical properties of least-squares estimators when using stochastic signals. The expected bias of the estimate caused by noisy measurement signals and Poisson moment functions is analyzed in detail. It requires knowledge from the disciplines of control engineering, stochastics, and signal processing. Consequently, a particular form of stochastics is necessary, which is summarized here.

## A.1 Random Variables

### A.1.1 Introduction

The concept of random variables as a numerical result of a random experiment was proved in the years around 1930 and made it possible to represent general random events in sets of real-valued numbers. The random variable is named by the capital letter, e.g., $X$. Realizations of the RV $X$ are abbreviated with the corresponding lower case letter, e.g., $x$.

According to [117], consider a random experiment with sample space or *domain* $\Omega$ and a single sample point $\omega \in \Omega$. A random variable (RV) is a real function that assigns a real number to each individual element $\omega \in \Omega$. The collection of all possible values of the RV is called *range*.

### A.1.2 Distribution Functions

The cumulative distribution function (CDF) is defined by

$$F_X(x) = P(X \leq x) \tag{A.1}$$

where $F_X$ denotes the probability that the RV $X$ takes a value less or equal $x \in [-\infty, \infty]$. Some of the most important properties include:

(i) cumulative distribution function (CDF) $F_X(x)$ is a non-decreasing function, i.e., it can increase or stay constant

(ii) the minimum value of CDF is zero, i.e., $F_X(-\infty) = 0$

(iii) the maximum value of CDF is one, i.e., $F_X(\infty) = 1$

(iv) $0 \leq F_X(x) \leq 1$

(v) probability interval $P(a < X \leq b) = F_X(b) - F_X(a)$

(vi) counter probability $P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$

### A.1.3 Continuous Random Variables

A RV $X$ is defined to be a continuous random variable (CRV) if there exists a non-negative PDF $f_X(x)\,,\ x \in [-\infty, \infty]$

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)\ . \tag{A.2}$$

Its properties read:

(i) $f_X(x)$ is non-negative, i.e., $f_X(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$

(iii) $P(a \leq X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x$, i.e. $P(x = a) = \int_a^a f_X(x)\,\mathrm{d}x = 0$

(iv) $P(X < a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x$

### A.1.4 Discrete Random Variables

The probability mass function (PMF) $p_X(x)$ of a discrete random variable (DRV) $X$ is defined by

$$p_X(x) = P(X = x) \tag{A.3}$$

with

$$\sum_{x=-\infty}^{\infty} p_X(x) = 1\ . \tag{A.4}$$

The CDF is a sequence of step functions and is defined by

$$F_X(x) = \sum_{k \leq x} p_X(k)\ . \tag{A.5}$$

## A.2 Moments of Random Variables

When dealing with random variables, it is common to consider the central tendency of the available data. The (weighted) arithmetic average or mean value is such a representative value. Assuming the data sorted by magnitude, the mean value is in the middle of this data set. Another parameter that characterizes the central tendency of data is called variance. It describes the spread of the data relative to its arithmetic average. The $n$-th moment of the random variable is defined by

$$\mathrm{E}[X^n] = \overline{X^n}\,, \quad n = 1, 2, 3, \ldots\,. \tag{A.6}$$

### A.2.1 Expectation

The first moment $\mathrm{E}[X]$, is the expected value (mean value) of a random variable $X$ and is defined by

$$\mathrm{E}[X] = \bar{X} = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \tag{A.7}$$

for CRVs and

$$\mathrm{E}[X] = \bar{X} = \sum_k x_k p_X(x_k) \tag{A.8}$$

for discrete random variables (DRVs).

### A.2.2 Conditional Expectation

The conditional expectation of the random variable $X$, assuming that an event A has occurred, is given by

$$\mathrm{E}[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x|A)\,\mathrm{d}x \tag{A.9}$$

and

$$\mathrm{E}[X|A] = \sum_k x_k p_{X|A}(x_k|A)\ . \tag{A.10}$$

The corresponding conditional probability density function (cond. PDF) $f_{X|A}(x|A)$ and conditional probability mass function (cond. PMF) $p_{X|A}(x|A)$ are defined by

$$\begin{aligned} f_{X|A}(x|A) &= \frac{f_X(x)}{P(A)}\,, \quad P(A) > 0 \\ p_{X|A}(x|A) &= \frac{p_X(x)}{P(A)}\,, \quad P(A) > 0 \end{aligned} \tag{A.11}$$

where $P(A)$ denotes the probability that an event $A$ occurs and $x \in A$.

### A.2.3 Variance and Higher Central Moments

In contrast to (A.6), the central moments are defined as moments of the difference between a random variable and its expected value, e.g., [117]. The $n$-th central moment reads

$$\mathrm{E}[\left(X-\bar{X}\right)^n] = \int_{-\infty}^{\infty} \left(x-\bar{X}\right)^n f_X(x)\,\mathrm{d}x \tag{A.12}$$

for CRVs and

$$\mathrm{E}[\left(X-\bar{X}\right)^n] = \sum_k \left(x_k-\bar{X}\right)^n p_X(x_k) \tag{A.13}$$

for DRVs.

The second-order central moment, i.e., $n=2$, is called variance $\sigma_X^2$ and obtained by

$$\sigma_X^2 = \mathrm{E}[\left(X-\bar{X}\right)^2] = \int_{-\infty}^{\infty} \left(x-\bar{X}\right)^2 f_X(x)\,\mathrm{d}x \tag{A.14}$$

or

$$\sigma_X^2 = \mathrm{E}[\left(X-\bar{X}\right)^2] = \sum_k \left(x_k-\bar{X}\right)^2 p_X(x_k) \tag{A.15}$$

respectively. Assuming $\bar{X}$ to be constant, the term $\sigma_X^2$ simplifies to

$$\begin{aligned}\sigma_X^2 &= \mathrm{E}[\left(X-\bar{X}\right)^2] = \mathrm{E}[X^2 - 2X\bar{X} + \bar{X}^2] \\ &= \mathrm{E}[X^2] - 2\bar{X}\,\mathrm{E}[X] + \bar{X}^2 \\ &= \mathrm{E}[X^2] - \bar{X}^2\,.\end{aligned} \tag{A.16}$$

## A.3 Stochastic Processes

### A.3.1 Continous-Time Stochastic Processes

A scalar function $x(t)$ is called a stochastic function or random function if the resulting value of $x$ at the time instance $t$ is only determined in a statistical sense. In other words, $x(t)$ is a sequence of random variables to consider the evolution of a random phenomenon concerning time, e.g., [126]. If a random experiment is repeatedly performed on the same conditions, these measurements $\{x_1(t), x_2(t), \ldots\, x_n(t)\}$ set a form of stochastic (random) time functions. The population of all possible time functions is termed *ensemble* of the random experiment or stochastic process $\mathrm{x}(t)$. Therefore, a stochastic process is the source of a random function.[1] An individual

[1]Actually, the notation of the stochastic process (random variable) must be different from that of the realization (possible result), e.g., $f_{\mathrm{x}}$ vs. $f_x$. For the sake of simplicity, often this distinction is not made when considering signals.

random time function $x(t)$ is denoted as a realization of the stochastic process $\mathrm{x}(t)$. To describe a stochastic process $\mathrm{x}(t)$, the CDF

$$F_{\mathrm{x}}(x,t) = P_{\mathrm{x}}(\mathrm{x}(t) \leq x) \tag{A.17}$$

or PDF

$$f_{\mathrm{x}}(x,t) = \frac{\mathrm{d}F_{\mathrm{x}}(x,t)}{\mathrm{d}x} \tag{A.18}$$

are necessary. The corresponding joint distributions explain internal coherence. Exemplarily, for $n$ points in time $\{t_1, t_2, \ldots, t_n\}$, the $n$-dimensional joint probability density function (joint PDF) read

$$f_{\mathrm{x}}(x_1, \ldots, x_n; t_1, \ldots, t_n) \tag{A.19}$$

and joint cumulative distribution function (joint CDF) is

$$F_{\mathrm{x}}(x_1, \ldots, x_n; t_1, \ldots, t_n) \ . \tag{A.20}$$

As stated in, e.g., [2, 3, 115], a stochastic process is fully characterized if the CDF and all joint CDFs

$$\begin{aligned} F_{\mathrm{x}}(x,t) &= P_{\mathrm{x}}(\mathrm{x}(t) \leq x) \ , \\ F_{\mathrm{x}}(x_1, \ldots, x_n; t_1, \ldots, t_n) &= P_{\mathrm{x}}(\mathrm{x}(t_1) \leq x_1, \ldots, \mathrm{x}(t_n) \leq x_n) \end{aligned} \tag{A.21}$$

or density functions

$$\begin{aligned} f_{\mathrm{x}}(x,t) &= \frac{\mathrm{d}F_{\mathrm{x}}(x,t)}{\mathrm{d}x} \ , \\ f_{\mathrm{x}}(x_1, \ldots, x_n; t_1, \ldots, t_n) &= \frac{\mathrm{d}^n F_{\mathrm{x}}(x_1, \ldots, x_n; t_1, \ldots, t_n)}{\mathrm{d}x_1 \ldots \mathrm{d}x_n} \end{aligned} \tag{A.22}$$

are known for $N \to \infty$. However, for the characterization of stochastic processes, it is usually sufficient to present all observations by using statistical averages, e.g., mean values or ACFs.

(i) **Expected value (mean or ensemble average)**

$$\bar{\mathrm{x}}(t) = \mathrm{E}[\mathrm{x}(t)] = \int_{-\infty}^{\infty} x f_{\mathrm{x}}(x,t)\, \mathrm{d}x \tag{A.23}$$

(ii) **Quadratic mean (variance)**

$$\sigma_x^2 = c_{\mathrm{xx}}(t,t) = \mathrm{E}[(\mathrm{x}(t) - \bar{\mathrm{x}}(t))^2] \tag{A.24}$$

(iii) **Autocovariance function**

$$c_{\mathrm{xx}}(t_1,t_2) = \mathrm{E}[(\mathrm{x}(t_1) - \bar{\mathrm{x}}(t_1))(\mathrm{x}(t_2) - \bar{\mathrm{x}}(t_2))] \tag{A.25}$$

(iv) **Autocorrelation function**

$$r_{\mathrm{xx}}(t_1,t_2) = \mathrm{E}[\mathrm{x}(t_1)\,\mathrm{x}(t_2)] = c_{\mathrm{xx}}(t_1,t_2) + \bar{\mathrm{x}}(t_1)\,\bar{\mathrm{x}}(t_2) \tag{A.26}$$

(v) **Crosscovariance function**

$$c_{\mathrm{xy}}(t_1,t_2) = \mathrm{E}[(\mathrm{x}(t_1) - \bar{\mathrm{x}}(t_1))(\mathrm{y}(t_2) - \bar{\mathrm{y}}(t_2))] \tag{A.27}$$

(vi) **Crosscorrelation function**

$$r_{\mathrm{xy}}(t_1,t_2) = \mathrm{E}[\mathrm{x}(t_1)\,\mathrm{y}(t_2)] = c_{\mathrm{xy}}(t_1,t_2) + \bar{\mathrm{x}}(t_1)\,\bar{\mathrm{y}}(t_2) \tag{A.28}$$

So far, it has been assumed that all distributions are functions of time. Hence, the stochastic process is termed *non-stationary*. However, it has not proven necessary to use such a broad, all-encompassing definition for many areas of application. Therefore, only certain classes of stochastic processes are treated in the following.

## A.3.2 Stationary Stochastic Processes

In the analysis of sequences or time series, however, the general theory of stochastic processes is not applicable. For this reason, it is usually assumed that stochastic processes are stationary and ergodic.

(i) **Stationarity** A random process is defined to be *strict sense stationary (SSS)* if the CDF and all joint CDFs are unaffected by a shift in time $\tau$

$$\begin{aligned} F_{\mathrm{x}}(x,t) &= F_{\mathrm{x}}(x,t+\tau)\ , \\ F_{\mathrm{x}}(x_1,\ldots,x_n;t_1,\ldots,t_n) &= F_{\mathrm{x}}(x_1,\ldots,x_n;t_1+\tau,\ldots,t_n+\tau) \end{aligned} \tag{A.29}$$

for any $\tau$, e.g., [127]. For the first-order distribution applies

$$\begin{aligned} F_{\mathrm{x}}(x;t) &= F_{\mathrm{x}}(x;t+\tau) = F_{\mathrm{x}}(x) \ , \\ f_{\mathrm{x}}(x;t) &= f_{\mathrm{x}}(x;t+\tau) = f_{\mathrm{x}}(x) \ . \end{aligned} \tag{A.30}$$

The second-order distribution only depends on the time difference $\tau = t_2 - t_1$, i.e.,

$$\begin{aligned} F_{\mathrm{x}}(x_1,x_2;t_1,t_2) &= F_{\mathrm{x}}(x_1,x_2;\tau) \ , \\ f_{\mathrm{x}}(x_1,x_2;t_1,t_2) &= f_{\mathrm{x}}(x_1,x_2;\tau) \ . \end{aligned} \tag{A.31}$$

As stated in [117], a random process is *SSS* if the ACF and autocovariance function do not depend on the time $t$, i.e.,

$$\begin{aligned} \bar{\mathrm{x}}(t) = \mathrm{E}[\mathrm{x}(t)] &= \int_{-\infty}^{\infty} \mathrm{x}(t) f_{\mathrm{x}}(x)\,\mathrm{d}x = \mathrm{const} \\ r_{\mathrm{xx}}(t,t+\tau) &= \mathrm{E}[\mathrm{x}(t)\,\mathrm{x}(t+\tau)] = r_{\mathrm{xx}}(\tau) \\ c_{\mathrm{xx}}(t,t+\tau) &= \mathrm{E}[(\mathrm{x}(t)-\bar{\mathrm{x}}(t))(\mathrm{x}(t+\tau)-\bar{\mathrm{x}}(t+\tau))] = c_{\mathrm{xx}}(\tau) \ . \end{aligned} \tag{A.32}$$

For $\tau = 0$, it applies

$$r_{\mathrm{xx}}(0) = \mathrm{E}[(x(t) - \bar{x}(t))] = \sigma_x^2 = \mathrm{const}\ . \tag{A.33}$$

In many practical tasks, only the mean value and the autocorrelation function of a random process are relevant. Hence, a weaker definition is introduced. If the stationarity condition (A.29) does not hold for all $n$ but holds for $n \leq 2$, a random process is called *wide (weak) sense stationary (WSS)*. It applies

$$\begin{aligned} \bar{\mathrm{x}}(t) = \mathrm{E}[\mathrm{x}(t)] &= \bar{\mathrm{x}} = \mathrm{const} \\ r_{\mathrm{xx}}(t,t+\tau) = \mathrm{E}[\mathrm{x}(t)\,\mathrm{x}(t+\tau)] &= r_{\mathrm{xx}}(\tau) \ . \end{aligned} \tag{A.34}$$

It is worth mentioning that a linear combination of stationary processes is also stationary [127].

(ii) **Ergodicity** A stationary process is called *ergodic* if the *ensemble average value* $\bar{\mathrm{x}}(t)$ is equal to the *temporal (time) average value*

$$\bar{x} = \overline{x(t)} = \mathrm{E}[x(t)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\,\mathrm{d}t \tag{A.35}$$

of each realization. For the quadratic mean or variance, it applies

$$\sigma_x^2 = \mathrm{E}[(x(t) - \bar{x})^2] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (x(t) - \bar{x})^2 \,\mathrm{d}t\,. \tag{A.36}$$

Ergodicity is (almost) always assumed in physical applications, e.g., for the analysis of time series. An ergodic process is always stationary. The opposite may not be valid.[2]

### A.3.3 Autocorrelation and Crosscorrelation

One can use the autocorrelation function (ACF) to describe internal coherences of the random function $x(t)$ with a time-shifted version $x(t+\tau)$ of itself. Thus, the ACF provides information about the underlying stochastic process $\mathrm{x}(t)$.

Let us assume an ergodic, i.e., stochastic process $\mathrm{x}(t)$ with a single realization $x(t)$. Then, the ACF reads

$$\begin{aligned} r_{\mathrm{xx}}(\tau) = \mathrm{E}[(x(t)\,x(t+\tau))] &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\,x(t+\tau)\,\mathrm{d}t \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t-\tau)\,x(t)\,\mathrm{d}t\,. \end{aligned} \tag{A.37}$$

For $\tau = 0$, the coherence

$$r_{\mathrm{xx}}(0) = \mathrm{E}[x^2(t)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)\,\mathrm{d}t \ < \infty \tag{A.38}$$

is a maximum and corresponds to the average signal power of $x(t)$. Consequently, it applies

$$r_{\mathrm{xx}}(\tau) \le r_{\mathrm{xx}}(0)\ . \tag{A.39}$$

The larger the time shift $\tau$ of the stochastic signals, the smaller the mutual dependence. For $\tau \to \infty$ the signals $x(t)$ and $x(t+\tau)$ are statistically independent, i.e. uncorrelated, resulting in

$$\lim_{\tau\to\infty} r_{\mathrm{xx}}(\tau) = r_{\mathrm{xx}}(\infty) = \lim_{\tau\to\infty} \left( \underbrace{\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\,\mathrm{d}t}_{=\mathrm{E}[x(t)]=\bar{x}} \underbrace{\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)\,\mathrm{d}t}_{=\mathrm{E}[x(t)]=\bar{x}} \right) = \bar{x}^2\,. \tag{A.40}$$

[2]A stochastic process is called *weak ergodic* if the compliance is only fulfilled for $\mathrm{E}[\mathrm{x}(t)]$ and $r_{\mathrm{xx}}(\tau)$.

Assuming a zero-mean stochastic process, it applies

$$\lim_{\tau \to \infty} r_{\mathrm{xx}}(\tau) = 0\,. \tag{A.41}$$

The main properties of the autocorrelation function of ergodic, i.e., stationary, processes read:

(i) the ACF is an even function, i.e $r_{\mathrm{xx}}(\tau) = r_{\mathrm{xx}}(-\tau)$,

(ii) for $\tau = 0$, the ACF $r_{\mathrm{xx}}(0) = \bar{x}^2$ denotes the average signal power,

(iii) for $\tau \to \infty$, $r_{\mathrm{xx}}(\infty) = \bar{x}^2$, means that the signals are uncorrelated,

(iv) $r_{\mathrm{xx}}(\tau) \leq r_{\mathrm{xx}}(0)$.

To describe the statistical coherence between different ergodic signals $x(t)$ and $y(t)$ the crosscorrelation function (CCF)

$$\begin{aligned} r_{\mathrm{xy}}(\tau) = \mathrm{E}[(x(t)\,y(t+\tau))] &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)\,y(t+\tau)\,\mathrm{d}t \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t-\tau)\,y(t)\,\mathrm{d}t \end{aligned} \tag{A.42}$$

is used with its statistical properties:

(i) the CCF is not an even function, i.e $r_{\mathrm{xy}}(\tau) = r_{\mathrm{yx}}(-\tau)$,

(ii) for $\tau = 0$, the CCF, $r_{\mathrm{xy}}(0) = \overline{x(t)\,y(t)}$ denotes mean of the product,

(iii) for $\tau \to \infty$, $r_{\mathrm{xy}}(\infty) = \bar{x}\bar{y}$, means the product of the mean values,

(iv) $r_{\mathrm{xy}}(\tau) \leq \frac{1}{2}(r_{\mathrm{x}}x(0) + r_{\mathrm{yy}}(0))$.

## A.3.4 Derivative Stochastic Processes

In [119, 120] a stochastic process $\mathrm{x}(t)$ or its realization (sample function) $x(t)$ and its related derivative $\dot{x}(t)$ is defined in mean square sense by

$$\lim_{\varepsilon\to 0} \mathrm{E}[\left|\frac{x(t+\varepsilon) - x(t)}{N\varepsilon} - \dot{x}(t)\right|^2]\,. \tag{A.43}$$

Let us assume a stationary stochastic function $x(t)$ and its first derivative $\dot{x}(t)$ with the corresponding ACFs and CCF

$$\begin{aligned} r_{xx}(\tau) &= r_{x^{(0)}x^{(0)}}(\tau) = \mathrm{E}[x(t)\,x(t+\tau)] = \mathrm{E}[x(t-\tau)\,x(t)] = r_{xx}^{(0)}(\tau) \\ r_{x\dot{x}}(\tau) &= r_{x^{(0)}x^{(1)}}(\tau) = \mathrm{E}[x(t)\,\dot{x}(t+\tau)] = \mathrm{E}[x(t-\tau)\,\dot{x}(t)] = r_{xx}^{(1)}(\tau) \\ r_{\dot{x}\dot{x}}(\tau) &= r_{x^{(1)}x^{(1)}}(\tau) = \mathrm{E}[\dot{x}(t)\,\dot{x}(t+\tau)] = \mathrm{E}[\dot{x}(t-\tau)\,\dot{x}(t)] = r_{xx}^{(2)}(\tau) \end{aligned} \tag{A.44}$$

and the definition for the $i$-th derivative function

$$r_{xx}^{(i)}(\tau) = \frac{\mathrm{d}^i}{\mathrm{d}\tau^i} r_{xx}(\tau)\ . \tag{A.45}$$

For the first derivative applies

$$r_{xx}^{(1)}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{E}[x(t)\,x(t+\tau)] = \mathrm{E}[x(t)\,\dot{x}(t+\tau)] = r_{x\dot{x}}(\tau)\ , \tag{A.46}$$

equivalent to

$$r_{xx}^{(1)}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{E}[x(t-\tau)\,x(t)] = -\,\mathrm{E}[\dot{x}(t-\tau)\,x(t)] = -r_{\dot{x}x}(\tau)\ . \tag{A.47}$$

From $r_{x\dot{x}}(\tau) = -r_{\dot{x}x}(\tau)$ follows for zero lag ($\tau = 0$)

$$r_{xx}^{(1)}(0) = r_{x\dot{x}}(0) = -r_{\dot{x}x}(0) = 0\,. \tag{A.48}$$

The second order derivative read

$$r_{xx}^{(2)}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{E}[x(t)\,\dot{x}(t+\tau)] = \mathrm{E}[x(t)\,\ddot{x}(t+\tau)] = r_{x\ddot{x}}(\tau) \tag{A.49}$$

and

$$r_{xx}^{(2)}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{E}[x(t-\tau)\,\dot{x}(t)] = -\,\mathrm{E}[\dot{x}(t-\tau)\,\dot{x}(t)] = -r_{\dot{x}\dot{x}}(\tau) \tag{A.50}$$

Hence, at $\tau = 0$, we get

$$r_{xx}^{(2)}(0) = -\,\mathrm{E}[\dot{x}^2(t)]\,. \tag{A.51}$$

Higher order derivatives can be determined iteratively. For example, the third derivation

$$r_{xx}^{(3)}(\tau) = r_{x^{(2)}x^{(1)}}(\tau) = -r_{x^{(3)}x^{(0)}}(\tau) = -r_{x^{(1)}x^{(2)}}(\tau) \tag{A.52}$$

with

$$r_{xx}^{(3)}(0) = r_{x^{(2)}x^{(1)}}(0) = -r_{x^{(3)}x^{(0)}}(0) = -r_{x^{(1)}x^{(2)}}(0) = 0 \tag{A.53}$$

or the fourth derivative

$$r_{xx}^{(4)}(\tau) = r_{x^{(2)}x^{(2)}}(\tau) = -r_{x^{(3)}x^{(1)}}(\tau) = -r_{x^{(1)}x^{(3)}}(\tau) = r_{x^{(4)}x^{(0)}}(\tau) \tag{A.54}$$

with

$$r_{xx}^{(4)}(0) = \mathrm{E}[\left(x^{(2)}(t)\right)^2] \, . \tag{A.55}$$

More general, for the ACF of a stationary random process $x(t)$ and associative time derivatives $x^{(i)}(t)$ or $x^{(j)}(t)$ applies

$$r_{x^{(j)}x^{(i)}}(\tau) = \mathrm{E}[x^{(j)}(t)\, x^{(i)}(t+\tau)] = (-1)^j r_{xx}^{(j+i)}(\tau) = (-1)^j \frac{\mathrm{d}^{j+i}}{\mathrm{d}\tau^{j+i}} r_{xx}(\tau) \, . \tag{A.56}$$

For $\tau = 0$:

$$r_{x^{(j)}x^{(i)}}(0) = \begin{cases} (-1)^{\frac{j+i}{2}} \mathrm{E}[\left(x^{\left(\frac{j+i}{2}\right)}(t)\right)^2] \, , & j+i = \{0,2,4,\dots\} \\ 0 \, , & j+i = \{1,3,5,\dots\} \end{cases} \tag{A.57}$$

## A.3.5 Discrete-Time Ergodic (Stationary) Stochastic Processes

In the field of control engineering, discrete-time (DT) stochastic signals usually result from the sampling of continuous-time (CT) stochastic signals at specific discrete time instances $t(k) = t_k = kT, 1 \leq k \leq N$ with sample time $T$. Since the statistical properties are very similar to those of the CT stochastic processes, and ergodic processes are assumed for practical applications, we settle for the most relevant results for stationary processes. A more detailed treatment of DT stochastic processes is given in, e.g., [128].

(i) **Expected value (mean value)**

$$\bar{x} = \mathrm{E}[x(k)] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x(k) \tag{A.58}$$

(ii) **Quadratic mean (variance)**

$$\sigma_x^2 = \mathrm{E}[(x(k) - \bar{x})^2] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (x(k) - \bar{x})^2 \tag{A.59}$$

(iii) **Autocovariance function**

$$\begin{aligned} c_{\mathrm{xx}}(\tau) = c_{\mathrm{xx};\tau} &= \mathrm{cov}(x,\tau) = \mathrm{E}[(x(k)-\bar{x})(x(k+\tau)-\bar{x})] \\ &= \mathrm{E}[x(k)\,x(k+\tau)] - \bar{x}^2 \end{aligned} \tag{A.60}$$

(iv) **Autocorrelation function**

$$r_{\mathrm{xx}}(\tau) = r_{\mathrm{xx};\tau} = \mathrm{E}[x(k)\,x(k+\tau)] = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} x(k)\,x(k+\tau) \tag{A.61}$$

(v) **Crosscovariance function**

$$\begin{aligned} c_{\mathrm{xy}}(\tau) = c_{\mathrm{xy};\tau} &= \mathrm{cov}(x,y,\tau) = \mathrm{E}[(x(k)-\bar{x})(y(k+\tau)-\bar{y})] \\ &= \mathrm{E}[x(k)\,y(k+\tau)] - \bar{x}\bar{y} \end{aligned} \tag{A.62}$$

(vi) **Crosscorrelation function**

$$\begin{aligned} r_{\mathrm{xy}}(\tau) = r_{\mathrm{xy};\tau} = \mathrm{E}[x(k)\,y(k+\tau)] &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} x(k)\,y(k+\tau) \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} x(k-\tau)\,y(k) \end{aligned} \tag{A.63}$$

### A.3.6 Stochastic Convergence Concepts

In contrast to the deterministic concept of convergence, several different notions of convergence of random variables exist in probability theory. The convergence of random variables to a limiting random variable is an essential concept in probability theory and its applications in statistics and stochastic processes. Hence, we introduce the most common concepts here: convergence in distribution, convergence in probability, convergence almost surely, and convergence in r-mean. The interested reader can find more information about stochastic convergence in, e.g., [109, 129]. In this thesis, the term convergence in probability is necessary to compute the estimation bias. Hence, we will look at this convergence concept a bit more closely.

The basic idea behind this notion of convergence is that the probability of an "unusual" result becomes smaller and smaller as the sequence progresses.

**Convergence in Probability**

**Definition A.1.** Given a sequence of random variables $X_n = \{X_1, X_2, \ldots\}$. $X_n$ converges in probability to the random variable $X$ if

$$\lim_{N\to\infty} P\{\|X_n - X\| \geq \epsilon\} = 0\,, \quad \epsilon > 0$$

for all $\epsilon > 0$. One may write

$$\operatorname*{plim}_{n\to\infty} X_n = X$$

for convergence in probability, e.g., [109, 129].

One can derive applicable calculation rules from the definition above, see [109].

**Convergence in Probability: Calculation Rules**

**Corollary A.1.** Let $X_n$, $Y_n$ be univariate random variables with observations $\{X_1, X_2, \ldots\}$ and $\{Y_1, Y_2, \ldots\}$. Assuming $\operatorname{plim}_{n\to\infty} X_n = a$ and $\operatorname{plim}_{n\to\infty} Y_n = b$, then applies:

$$\begin{aligned} \operatorname*{plim}_{n\to\infty}(X_n \pm Y_n) &= \operatorname*{plim}_{n\to\infty}(X_n) \pm \operatorname*{plim}_{n\to\infty}(Y_n) = a \pm b\,, \\ \operatorname*{plim}_{n\to\infty}(X_n Y_n) &= \operatorname*{plim}_{n\to\infty}(X_n) \operatorname*{plim}_{n\to\infty}(Y_n) = ab\,, \\ \operatorname*{plim}_{n\to\infty}\left(\frac{X_n}{Y_n}\right) &= \frac{\operatorname{plim}_{n\to\infty}(X_n)}{\operatorname{plim}_{n\to\infty}(Y_n)} = \frac{a}{b}\,, \text{ if } b \neq 0\,. \end{aligned} \tag{A.64}$$

If $\mathbf{A}$ is an adequate and nonsingular matrix of random variables, then:

$$\operatorname*{plim}_{n\to\infty}\left(\mathbf{A}_n^{-1}\right) = \left(\operatorname*{plim}_{n\to\infty}(\mathbf{A}_n)\right)^{-1}\,. \tag{A.65}$$

## A.3.7 Gaussian White Noise Process

One denotes a DT random stationary and ergodic process $\mathrm{x}_k$ with a realization $x_k$ as white if it applies:

(i) the expectation of each element is zero, i.e., $\mathrm{E}[x_k] = 0$

(ii) the variance of each element is constant and finite, i.e., $\operatorname{var}(x_k) < \infty$

(iii) the elements are serially uncorrelated, i.e., $r_{xx;k} = 0$ for all $k \neq 0$

Additionally, suppose the random variables are independent, i.e.,

$$\operatorname{cov}\left(x_m, x_{m+k}\right) = 0 \tag{A.66}$$

for any $k \neq 0$. In that case, the random noise sequence is independent and identically distributed (i.i.d.), and one can call it *white Gaussian noise.* For white Gaussian noise applies

$$x_k \sim i.i.d\left(0, \sigma^2\right) \,. \tag{A.67}$$

Every white Gaussian noise sequence is (serially) uncorrelated, but the converse is not generally true. Please note that the continuous-time (CT) Gaussian white noise process variance is mathematically defined to be infinite, e.g. [130]. However, we obtain a DT process with a finite variance if the samples are obtained by sampling from a CT Gaussian white noise process with finite sample time $T$.

## A.4 Linear Time-Invariant Systems with Stochastic Input

In this section, we consider the effects of stochastic (random) input signals on the output of linear time-invariant (LTI) systems, treated in, e.g., [115–118]. For reasons of practical applications, we restrict ourselves to ergodic, i.e., stationary, input signals. All considerations are carried out on continuous-time (CT) signals or systems, respectively. However, one can apply the results to DT models.

The random input signal $x(t)$ is a realization of the underlying (ergodic) stochastic process $\mathrm{x}(t)$ and is characterized by its correlation function $r_{xx}(\tau)$. The LTI system is characterized by its transfer function $G(s)$ or impulse response $g(t)$. The model output $y(t)$ is also a realization of a random process, and it applies

$$y(t) = (x * g)(t) = \int_{-\infty}^{\infty} x(\tau)\, g(t-\tau)\, \mathrm{d}\tau = \int_{-\infty}^{\infty} g(\tau)\, x(t-\tau)\, \mathrm{d}\tau = (g * x)(t) \tag{A.68}$$

where ‘*’ denotes the convolution operator.

Inserting (A.68) in (A.35) gives

$$\bar{y} = \mathrm{E}[y(t)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t)\, \mathrm{d}t = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{-\infty}^{\infty} g(\tau)\, x(t-\tau)\, \mathrm{d}\tau\, \mathrm{d}t \,. \tag{A.69}$$

By interchanging the order of integration and shifting the limit, we obtain

$$\begin{aligned} \bar{y} = \mathrm{E}[y(t)] &= \int_{-\infty}^{\infty} g(\tau) \left( \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t-\tau)\,\mathrm{d}t \right) \mathrm{d}\tau \\ &= \mathrm{E}[x(t)] \int_{-\infty}^{\infty} g(\tau)\,\mathrm{d}\tau\,. \end{aligned} \tag{A.70}$$

For causal systems, the impulse response vanishes for $t < 0$, and we finally get

$$\mathrm{E}[y(t)] = \mathrm{E}[x(t)] \int_{0}^{\infty} g(\tau)\,\mathrm{d}\tau\,. \tag{A.71}$$

Consequently, a zero-mean stochastic input signal results in a zero-mean stochastic output signal.

The ACF for the model output read

$$r_{yy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t)\,y(t+\tau)\,\mathrm{d}t = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t-\tau)\,y(t)\,\mathrm{d}t\,. \tag{A.72}$$

As shown in [115], inserting (A.68), and introducing

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} g(u)\,x(t-u)\,\mathrm{d}u \\ y(t+\tau) &= \int_{-\infty}^{\infty} g(v)\,x(t+\tau-v)\,\mathrm{d}v \end{aligned} \tag{A.73}$$

results in

$$r_{yy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t-u)\,x(t+\tau-v)\,g(u)\,g(v)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}t\,. \tag{A.74}$$

Again, interchanging the order of integration and shifting the limit gives

$$r_{yy}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u)\,g(v) \left( \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t-u)\,x(t+\tau-v)\,\mathrm{d}t \right) \mathrm{d}u\,\mathrm{d}v \tag{A.75}$$

Substituting $t - u = w$, and hence $t + \tau - v = w + \tau + u - v$ gives

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(w)\,x(w+\tau+u-v)\,\mathrm{d}w = r_{xx}(\tau+u-v) \tag{A.76}$$

and (A.75) simplifies to

$$r_{yy}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u)\,g(v)\,r_{xx}(\tau+u-v)\,\mathrm{d}u\,\mathrm{d}v\,. \tag{A.77}$$

Substituting $v - u = \lambda$ results in

$$\begin{aligned} r_{yy}(\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)\, g(u+\lambda)\, r_{xx}(\tau-\lambda)\,\mathrm{d}u\,\mathrm{d}\lambda \\ &= \int_{-\infty}^{\infty} r_{xx}(\tau-\lambda)\left(\int_{-\infty}^{\infty} g(u)\, g(u+\lambda)\,\mathrm{d}u\right)\mathrm{d}\lambda\,. \end{aligned} \tag{A.78}$$

In contrast to stochastic signals, the ACF for a deterministic aperiodic function $g(t)$ is defined by

$$r_{gg}(\tau) = \int_{-\infty}^{\infty} g(t)\, g(t+\tau)\,\mathrm{d}t\,. \tag{A.79}$$

Inserting (A.79) in (A.78) finally results in

$$r_{yy}(\tau) = \int_{-\infty}^{\infty} r_{xx}(\tau-\lambda)\, r_{gg}(\lambda)\,\mathrm{d}\lambda = (r_{xx} * r_{gg})(\tau)\ . \tag{A.80}$$

To summarize, the ACF of the output is calculated by the convolution of the ACFs of the stochastic input signal $x(t)$ and the impulse response $g(t)$.

For DT signals or models, the expected value is

$$\mathrm{E}[y(t_k)] = \mathrm{E}[y_k] = \mathrm{E}[x_k]\sum_{k=1}^{\infty} g_k \tag{A.81}$$

and for the ACF applies

$$r_{yy}(k) = r_{yy,k} = (r_{xx} * r_{gg})_k = (r_{gg} * r_{xx})_k = \sum_{i=-\infty}^{\infty} r_{gg,i} r_{xx;i-k}\,. \tag{A.82}$$

## A.4.1 Filtered White Noise

Due to the present application case in Section 3.4, we limit ourselves to discrete-time (DT) signals. Consequently, we assume a DT ergodic, i.e., stationary, white noise process $\mathrm{e}_k$ with a realization $e_k$. A noise sequence $e_k$ is called *white* if successive samples are uncorrelated in time. It applies

$$r_{ee,k} = \mathrm{E}[e_m e_{m+k}] = \sigma_e^2\delta_k = \begin{cases}\sigma_e^2\,, & k = 0\\ 0\,, & k \neq 0\end{cases} \tag{A.83}$$

for any $m$ and constant noise variance $\sigma_e^2$.

By filtering white noise, the individual samples are no longer uncorrelated. Generally speaking, we obtain *colored noise*. Assuming a strictly stable LTI system (filter) with impulse response $g_k$, the resulting colored noise is stationary. Thus, the filtered white noise sequence $v_k$ reads

$$v_k = (e * g)_k = (g * e)_k = \sum_{i=-\infty}^{\infty} e_i g_{i-k} \, . \tag{A.84}$$

Convolving any signal $x_k$ with the delta function $\delta_k$ results in precisely the same signal

$$(\delta * x)_k = (x * \delta)_k = x_k \tag{A.85}$$

and makes the delta function the convolution identity. Finally, inserting (A.85) in (A.82) gives

$$r_{vv,k} = (r_{ee} * r_{gg})_k = \left(\sigma_e^2 \delta * r_{gg}\right)_k = \sigma_e^2 r_{gg,k} \, . \tag{A.86}$$

In other words, the ACF of filtered white noise equals the ACF of the filter's impulse response $g_k$ times the white-noise variance $\sigma_e^2$.

# Bibliography

[1] K. J. Keesman, *System identification: An introduction*, ser. Advanced textbooks in control and signal processing. London and New York: Springer, 2011.

[2] R. Isermann and M. Münchhof, *Identification of Dynamic Systems.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[3] C. Bohn and H. Unbehauen, *Identifikation dynamischer Systeme.* Wiesbaden: Springer Fachmedien Wiesbaden, 2016.

[4] R. A. Fisher, *The Design of Experiments*, 1st ed. Oliver and Boyd, 1935.

[5] D. Saha and G. Rao, *Identification of Continuous Dynamical Systems: The Poisson Moment Functional (PMF) Approach*, ser. Lecture Notes in Control and Information Sciences. Berlin Heidelberg: Springer-Verlag, 1983, no. 56.

[6] H. Garnier, P. Sibille, and T. Bastogne, "A bias-free least-squares parameter estimator for continuous-time state-space models," in *Proceedings of the 36th IEEE Conference on Decision and Control.* IEEE, 1997, pp. 1860–1865.

[7] K. Wada, Z.-J. Yang, and S. Sagara, "On the bias-compensated least-squares estimation for identification of continuous systems," in *IFAC Workshop on Automatic Control for Quality and Productivity (ACQP'92)*, ser. 8, vol. 25, 1992, pp. 455–462.

[8] K. Siebertz, D. van Bebber, and T. Hochkirchen, *Statistische Versuchsplanung*, 1st ed. Springer-Verlag Berlin Heidelberg, 2010.

[9] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, 1st ed. Academic Press, 1977, vol. 136.

[10] T. Söderström and P. Stoica, *System Identification*, ser. Series in Systems and Control Engineering. New York, NY: Prentice-Hall, 1989.

[11] L. Ljung, *System identification: Theory for the user*, 2nd ed., ser. Prentice Hall information and system sciences series. Upper Saddle River, NJ: Prentice Hall PTR, 2012.

[12] N. Tietze, “Model-based Calibration of Engine Control Units Using Gaussian Process Regression,” Dissertation, Technische Universität Darmstadt, Darmstadt, Germany, 2015.

[13] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin and Heidelberg: Springer, 2001.

[14] M. Deflorian and F. Klöpper, “Design of dynamic experiments,” in *Design of experiments (DoE) in engine development*, K. Röpke, Ed. Renningen: Expert Verl., 2009, vol. V.

[15] M. Deflorian and S. Zaglauer, “Design of Experiments for nonlinear dynamic system identification: 18th IFAC World Congress,” *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 13 179–13 184, 2011.

[16] A. Schreiber, M. Kowalczyk, and R. Isermann, “Method for Dynamic Online Identification with Integrated determination of Operating Boundaries,” in *Design of experiments (DoE) in engine development*, K. Röpke, Ed. Renningen: Expert Verl., 2011, vol. VI, pp. 120–135.

[17] C. P. Majer, “Parameterschätzung, versuchsplanung und trajektorienoptimierung für verfahrenstechnische prozesse,” Ph.D. dissertation, Universität Stuttgart, 1997.

[18] G. Franceschini and S. Macchietto, “Model-based design of experiments for parameter precision: State of the art,” *Chemical Engineering Science*, vol. 63, no. 19, pp. 4846–4872, 7 2008.

[19] G. C. Goodwin, “Experiment design,” in *6th IFAC Symposium on Identification and System Parameter Estimation*, ser. 6, vol. 15, no. 4, 1982, pp. 65–71.

[20] E. Walter and L. Pronzato, “Qualitative and quantitative experiment design for phenomenological model – a survey,” *Automatica*, vol. 26, pp. 195–213, 1990.

[21] É. Walter and L. Pronzato, *Identification of parametric models from experimental data*, ser. Communications and control engineering series. London: Springer,

1997. [Online]. Available: http://www.loc.gov/catdir/enhancements/fy0815/96049169-d.html

[22] G. E. P. Box and H. L. Lucas, "Design of experiments in non-linear situations," *Biometrika*, vol. 46, pp. 77–90, 1959.

[23] R. K. Mehra, "Optimal input signals for parameter estimation in dynamic systems–survey and new results," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 753–768, 12 1974.

[24] L. C. Zullo, "Computer aided design of experiments. an engineering approach." Ph.D. dissertation, University of London, 1991.

[25] G. C. Goodwin, *Identification: experiment design.* Pergamom Press, 1987, no. 4, pp. 2257–2264.

[26] D. De Pauw, "Optimal experimental design for calibration of bioprocess models: a validated software toolbox," Ph.D. dissertation, Ghent University, 2005.

[27] M. Caracotsios and W. E. Stewart, "Sensitivity analysis of initial value problems with mixed odes and algebraic equations," *Computers & Chemical Engineering*, vol. 9, no. 4, pp. 359–365, 1985. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0098135485850146

[28] J. A. Jacquez and P. Greif, "Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design," *Mathematical Biosciences*, vol. 77, no. 1-2, pp. 201–227, 1985.

[29] R. Schenkendorf, X. Xie, M. Rehbein, S. Scholl, and U. Krewer, "The Impact of Global Sensitivities and Design Measures in Model-Based Optimal Experimental Design," *Processes*, vol. 6, no. 4, p. 27, 2018.

[30] B. Iooss and P. Lemaître, "A Review on Global Sensitivity Analysis Methods," in *Uncertainty management in simulation-optimization of complex systems*, ser. Operations Research / Computer Science Interfaces Series, G. Dellino and C. Meloni, Eds. Boston, MA: Springer US, 2015, pp. 101–122.

[31] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Sensitivity analysis for chemical models," *Chemical Reviews*, vol. 105, no. 7, pp. 2811–2828, 2005.

[32] J. Zádor, I. Zsély, and T. Turányi, “Local and global uncertainty analysis of complex chemical kinetic systems,” *Reliability Engineering & System Safety*, vol. 91, no. 10-11, pp. 1232–1240, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832005002322

[33] Y. Chu and J. Hahn, “Necessary condition for applying experimental design criteria to global sensitivity analysis results,” *Computers & Chemical Engineering*, vol. 48, pp. 280–292, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098135412002852

[34] M. Rodriguez-Fernandez, S. Kucherenko, C. Pantelides, and N. Shah, “Optimal experimental design based on global sensitivity analysis,” in *Computer Aided Chemical Engineering : 17 European Symposium on Computer Aided Process Engineering*, V. Pleşu and P. Ş. Agachi, Eds. Elsevier, 2007, vol. 24, pp. 63–68. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570794607800344

[35] G. C. Goodwin, C. R. Rojas, and J. S. Welsh, “Good, Bad and Optimal Experiments for Identification,” in *Forever Ljung in System Identification*, T. Glad and G. Hendeby, Eds. Lund, Sweden: Studentlitteratur AB, 2006. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-72587

[36] G. P. Rao and H. Unbehauen, “Identification of continuous-time systems: IEE Proceedings - Control Theory and Applications, 153(2), 185-220,” *IEE Proceedings - Control Theory and Applications*, vol. 153, no. 2, pp. 185–220, 2006.

[37] H. Unbehauen and G. Rao, “A review of identification in continuous systems – a survey,” *Automatica*, vol. 26, no. 1, pp. 23–35, 1990.

[38] ——, “A review of identification in continuous systems,” *IFAC Annual Reviews in Control*, vol. 22, pp. 145–171, 1998.

[39] K. J. Åström and P. Eykhoff, “System identification—A survey: Automatica, 7(2), 123-162,” *Automatica*, vol. 7, no. 2, pp. 123–162, 1971.

[40] P. C. Young, “Parameter estimation of continuous-time models – a survey,” *Automatica*, vol. 17, no. 2, pp. 23–39, 1981.

[41] Y. Tomita, A. A. H. Damen, and P. M. J. van den Hof, “Difference between Equation Error Method (EEM) and Output Error Method (OEM)

in System Identification," *Transactions of the Society of Instrument and Control Engineers*, vol. 22, no. 1, pp. 50–55, 1986. [Online]. Available: https://www.jstage.jst.go.jp/article/sicetr1965/22/1/22_1_50/_article

[42] ——, "Equation error versus output error methods," *Ergonomics*, vol. 35, no. 5-6, pp. 551–564, 1992.

[43] J. P. Norton, *An introduction to identification*, 2nd ed. London: Acad. Press, 1988.

[44] R. C. K. Lee, *Optimal estimation, identification, and control*, 3rd ed., ser. Massachusetts Institute of Technology (Cambridge Mass.). Research monograph. Cambridge Mass.: M. I. T. P, 1964, vol. No. 28.

[45] A. Albert and R. W. Sittler, "A Method for Computing Least Squares Estimators that Keep Up with the Data," *Journal of the Society for Industrial and Applied Mathematics, Series A Control*, vol. 3, no. 3, pp. 384–417, 1966.

[46] R. Deutsch, *Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1966.

[47] P. Eykhoff, *System identification: Parameter and state estimation*. Chichester, England: Wiley, 1974.

[48] K. Steiglitz and L. McBride, "A technique for the identification of linear systems," *IEEE Transactions on Automatic Control*, vol. 10, no. 4, pp. 461–464, 1965.

[49] R. Isermann, *Prozeßidentifikation: Identifikation und Parameterschätzung dynamischer Prozesse mit diskreten Signalen*, ser. Hochschultext. Berlin and Heidelberg: Springer, 1974.

[50] P. Stoica and T. Soderstrom, "A method for the identification of linear systems using the generalized least squares principle," *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 631–634, 1977.

[51] R. Hastings-James and M. W. Sage, "Recursive generalised-least-squares procedure for online identification of process parameters: Proceedings of the Institution of Electrical Engineers, 116(12), 2057," *Proceedings of the Institution of Electrical Engineers*, vol. 116, no. 12, p. 2057, 1969.

[52] P. C. Young, "The use of linear regression and related procedures for the identification of dynamic processes," in *Seventh Symposium on Adaptive Processes*, 1968, p. 53.

[53] V. Panuska, "An adaptive recursive-least-squares identification algorithm," in *1969 IEEE Symposium on Adaptive Processes (8th) Decision and Control*, 1969, p. 65.

[54] G. H. Golub and C. F. van Loan, "An Analysis of the Total Least Squares Problem," *SIAM Journal on Numerical Analysis*, vol. 17, no. 6, pp. 883–893, 1980. [Online]. Available: http://www.jstor.org/stable/2156807

[55] S. van Huffel and J. Vandewalle, *The total least squares problem: Computational aspects and analysis*, ser. Frontiers in applied mathematics. Philadelphia, Pa: Society for Industrial and Applied Mathematics (SIAM 3600 Market Street Floor 6 Philadelphia PA 19104), 1991, vol. vol. 9. [Online]. Available: http://epubs.siam.org/ebooks/siam/frontiers_in_applied_mathematics/fr09

[56] I. Markovsky and S. van Huffel, "Overview of total least-squares methods," *Signal Processing*, vol. 87, no. 10, pp. 2283–2302, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168407001405

[57] B. Schaffrin, "A note on Constrained Total Least-Squares estimation," *Linear Algebra and its Applications*, vol. 417, no. 1, pp. 245–258, 2006.

[58] S. Sagara and K. Wada, "On-line modified least-squares parameter estimation of linear discrete dynamic systems," *International Journal of Control*, vol. 25, no. 3, pp. 329–343, 1977.

[59] M. Eguchi, K. Wada, and K. Nakano, "Identification of Pulse Transfer Function in the Presence of Input and Output Noise," *IFAC Proceedings Volumes*, vol. 25, no. 8, pp. 463–470, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667017540987

[60] Li-Juan Jia, M. Ikenoue, Chun-Zhi Jin, and K. Wada, "On bias compensated least squares method for noisy input-output system identification," in *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228)*, 2001, pp. 3332–3337 vol.4.

[61] M. Ikenoue, S. Kanae, Z.-J. Yang, and K. Wada, "Identification of noisy input-output system using bias-compensated least-squares method," *IFAC*

*Proceedings Volumes*, vol. 38, no. 1, pp. 803–808, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S147466701636147X

[62] P. C. Young, "Some observations on instrumental variable methods of time-series analysis," *International Journal of Control*, vol. 23, no. 5, pp. 593–612, 1976.

[63] T. Söderström and P. G. Stoica, *Instrumental Variable Methods for System Identification*, ser. Lecture Notes in Control and Information Sciences. Berlin and Heidelberg: Springer, 1983, vol. 57.

[64] P. Young, "The Instrumental Variable Method: A Practical Approach to Identification and System Parameter Estimation," *IFAC Proceedings Volumes*, vol. 18, no. 5, pp. 1–15, 1985. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667017605354

[65] M. Gilson, "What has Instrumental Variable method to offer for system identification?" *IFAC-PapersOnLine*, vol. 48, no. 1, pp. 354–359, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896315001779

[66] P. Joseph, J. Lewis, and J. Tou, "Plant identification in the presence of disturbances and application to digital adaptive systems," *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, vol. 80, no. 1, pp. 18–24, 1961.

[67] P. C. Young, "An instrumental variable method for real-time identification of a noisy process," *Automatica*, vol. 6, no. 2, pp. 271–287, 1970. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005109870900981

[68] J. R. Raol, G. Girija, and J. Singh, *Modelling and Parameter Estimation of Dynamic Systems*, ser. IET control engineering series. London: IET, 2004, vol. 65.

[69] T. Söderström, "An On-line Algorithm for Approximate Maximum Likelihood Identification of Linear Dynamic Systems," Lund, Sweden.

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

[71] K. J. Åström, P. Hagander, and J. Sternby, "Zeros of sampled systems," *Automatica*, vol. 20, no. 1, pp. 31–38, 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005109884900621

[72] R. H. Middleton and G. C. Goodwin, *Digital control and estimation: A unified approach*, ser. Prentice-Hall information and system sciences series. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[73] P. C. Young, A. Chotai, and W. Tych, "Identification, Estimation and Control of Continuous-Time Systems Described by Delta Operator Models," in *Identification of Continuous-Time Systems*, ser. International Series on Microprocessor-Based Systems Engineering, N. K. Sinha, G. P. Rao, and G. Prasada Rao, Eds. Dordrecht and Boston: Springer Science/Kluwer Academic, 1991, vol. 7.

[74] N. K. Sinha, G. P. Rao, and G. Prasada Rao, *Identification of Continuous-Time Systems: Methodology and Computer Implementation*, ser. International Series on Microprocessor-Based Systems Engineering. Dordrecht and Boston: Springer Science/Kluwer Academic, 1991, vol. 7. [Online]. Available: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10645317

[75] T. Söderström and M. Mossberg, "Performance evaluation of methods for identifying continuous-time autoregressive processes," *Automatica*, vol. 36, no. 1, pp. 53–59, 2000. [Online]. Available: https://www.researchgate.net/publication/223254507_Performance_evaluation_of_methods_for_identifying_continuous-time_autoregressive_processes

[76] O. Cieza Aguirre, J. C. Tafur, and J. Reger, "Frequency domain modulating functions for continuous-time identification of linear and nonlinear systems," in *Proceedings of 16th Latinamerican Control Conferenc*, 10 2014, pp. 690–695.

[77] M. Shinbrot, "On the analysis of linear and nonlinear system," *Transaction ASME*, vol. 79, pp. 547–552, 1957.

[78] W. Byrski, S. Fuksa, and M. Nowak, "The quality of identification for different normalizations of continuous transfer functions," in *Proceedings of the 22nd IASTED International Conference on Modelling, Identification, and Control (MIC 2003)*, M. H. Hamza, Ed. ACTA Press, 2003.

[79] J. Jouffroy and J. Reger, "Finite-time simultaneous parameter and state estimation using modulating functions," in *2015 IEEE Conference on Control Applications (CCA).* Piscataway, NJ: IEEE, 2015, pp. 394–399.

[80] J. Loeb and G. Cahen, "More about process identification," *IEEE Transactions on Automatic Control*, vol. 10, no. 3, pp. 359–361, 1965.

[81] A. E. Pearson, "Explicit parameter identification for a class of nonlinear input/output differential operator models," in *Proceedings of the 31st IEEE Conference on Decision and Control*, 1992, pp. 3656–3660 vol.4.

[82] S. A. Jalali, J. R. Jordan, and R. Mackie, "Measurement of the parameters of all-pole transfer functions using shifted hermite modulating functions," *Automatica*, vol. 28, no. 3, pp. 613–616, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/000510989290186J

[83] A. E. Pearson and F. C. Lee, "Time limited identification of continuous systems using trigonometric modulation functions," in *The Third Yale Workshop on Applications of Adaptive Systems Theory*, Yale University, Ed., 1983, pp. 168–173.

[84] S. Daniel-Berhe and H. Unbehauen, "Parameter estimation of nonlinear continuous-time systems using Hartley modulating functions," in *Proceedings of UKACC International Conference On Control'96.* Shrewsbury: Field Studies Council, 1996, pp. 228–233 vol.1.

[85] A. Patra and H. Unbehauen, "Identification of a class of nonlinear continuous-time systems using Hartley modulating functions," *International Journal of Control*, vol. 62, no. 6, pp. 1431–1451, 1995.

[86] T. Janiczek, "Generalization of the modulating functions method into the fractional differential equations," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 58, no. 4, 2010.

[87] S. Asiri, D.-Y. Liu, and T.-M. Laleg-Kirati, "Selection of Modulating Functions' Design Parameters for Estimation Problems," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 277–282, 2021.

[88] D. C. Saha and G. P. Rao, *Identification of Continuous Dynamical Systems: The Poisson Moment Functional (PMF) Approach*, ser. Lecture Notes in Control and Information Sciences. Berlin and Heidelberg: Springer, 1983, vol. 56.

[89] D. Saha and G. P. Rao, “A general algorithm for parameter identification in lumped continuous systems–The Poisson moment functional approach,” *IEEE Transactions on Automatic Control*, vol. 27, no. 1, pp. 223–225, 1982.

[90] D. C. Saha and G. P. Rao, “Identification of lumped linear systems in the presence of unknown initial conditions via Poisson moment functionals,” *International Journal of Control*, vol. 31, no. 4, pp. 637–644, 1980.

[91] J. E. Diamessis, “A new method for determining the parameters of physical systems,” *Proceedings of the IEEE*, vol. 53, no. 2, pp. 205–206, 1965.

[92] ——, “On the determination of the parameters of certain nonlinear systems,” *Proceedings of the IEEE*, vol. 53, no. 5, p. 556, 1965.

[93] ——, “A method for determining the parameters of certain time-varying systems,” *Proceedings of the IEEE*, vol. 53, no. 4, pp. 396–397, 1965.

[94] S. Sagara and Z.-Y. Zhao, “Numerical Integration Approach to On-line Identification of Continuous Systems in Presence of Measurement Noise,” *IFAC Proceedings Volumes*, vol. 21, no. 9, pp. 441–446, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667017547679

[95] S. Sagara and Z.-Y. ZHAO, “Recursive identification of transfer function matrix in continuous systems via linear integral filter,” *International Journal of Control*, vol. 50, no. 2, pp. 457–477, 1989.

[96] S. Sagara, Z.-J. Yang, and K. Wada, “Identification of continuous systems using digital low-pass filters,” *International Journal of Systems Science*, vol. 22, no. 7, pp. 1159–1176, 1991.

[97] Z.-Y. ZHAO, S. Sagara, and M. TOMIZUKA, “A new bias-compensating LS method for continuous system identification in the presence of coloured noise,” *International Journal of Control*, vol. 56, no. 6, pp. 1441–1452, 1992.

[98] H. Garnier, P. Sibille, H. L. Nguyen, and T. Spott, “A Bias-Compensating Least-Squares Method for Continuous-Time System Identification Via Poisson Moment Functionals,” *IFAC Proceedings Volumes*, vol. 27, no. 8, pp. 1591–1596, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667017479389

[99] S. Mayr, G. Grabmair, and J. Reger, “Input design and online system identification based on Poisson moment functions for system outputs with quantization noise,” in *2017 25th Mediterranean Conference on Control and Automation (MED).* Piscataway, NJ: IEEE, 2017, pp. 23–29.

[100] P. C. Young, “Optimal IV Identification and Estimation of Continuous-Time TF Models,” *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 109–114, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667015394258

[101] J. Karlsson, M. Anguelova, and M. Jirstrand, “An Efficient Method for Structural Identifiability Analysis of Large Dynamic Systems,” *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 941–946, 2012.

[102] M. Noack, K. Wulff, J. Reger, and M. Höper, “Observability analysis and nonlinear observer design for a turbocharger in a diesel engine,” in *2016 IEEE Conference on Control Applications (CCA)*, 2016, pp. 323–328.

[103] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou, “Structural Identifiability Analysis via Extended Observability and Decomposition,” *IFAC-PapersOnLine*, vol. 49, no. 26, pp. 171–177, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896316327847

[104] A. F. Villaverde, “Observability and Structural Identifiability of Nonlinear Biological Systems,” *Complexity*, pp. 1–12, 2019.

[105] A. Isidori, *Nonlinear Control Systems*, 3rd ed., ser. Communications and Control Engineering. London: Springer, 1995.

[106] A. Kugi, “Nichtlineare Regelung elektrischer Systeme [in German] (Nonlinear control of electrical systems),” Dissertation, JKU Linz, 1995.

[107] S. Mayr, G. Grabmair, and J. Reger, “Fast Model-Based Fault Detection in Single-Phase Photovoltaic Systems,” in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society.* Piscataway, NJ: IEEE, 2019, pp. 4615–4622.

[108] B. Eitzinger, “Parameter Identification for Nonlinear Dynamical Systems,” Dissertation, JKU, Linz, Austria, 2002.

[109] M. J. Panik, *Advanced Statistics from an Elementary Point of View.* Amsterdam: Elsevier/Academic Press, 2006. [Online]. Available: http://www.loc.gov/catdir/enhancements/fy0622/2005009834-d.html

[110] R. Deutsch, *Estimation Theory.* Englewood Cliffs, N.J. : Prentice-Hall, 1965.

[111] N. E. Nahi, *Estimation Theory and Applications.* New York: Wiley, 1969.

[112] M. Fisz, *Wahrscheinlichkeitsrechnung und Mathematische Statistik*, 2nd ed. Berlin: VEB Deutscher Verlag der Wissenschaften, 1963.

[113] J. Rudolph, J. Birk, and M. Zeitz, "Dependence on Time Derivatives of the Input in the Nonlinear Controller Canonical Form," in *New Trends in Systems Theory*, ser. Progress in Systems and Control Theory, G. Conte, A. M. Perdon, and B. Wyman, Eds. Boston, MA: Birkhäuser Boston, 1991, pp. 636–643.

[114] J. H. L. White and J. S. Cho, "A Three Line Proof that OLS is BLUE," *SSRN Electronic Journal*, 2012. [Online]. Available: https://ssrn.com/abstract=2050611

[115] M. Meyer, *Signalverarbeitung*, 8th ed. Springer Vieweg, 2017.

[116] S. Kay, *Intuitive Probability and Random Processes using MATLAB(R).* Springer US, 2006.

[117] O. Ibe, *Fundamentals of Applied Probability and Random Processes.* Elsevier, 2014.

[118] H. W. Knobloch and H. Kwakernaak, *Lineare Kontrolltheorie.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1985.

[119] Z. Chen, S. Haykin, J. J. Eggermont, and S. Becker, *Correlative Learning.* Hoboken, NJ, USA: John Wiley & Sons, Inc, 2007.

[120] J. S. Bendat and A. G. Piersol, *Random data: Analysis and measurement procedures*, 4th ed., ser. Wiley series in probability and statistics. Hoboken, New Jersey: Wiley a John Wiley & Sons Inc, 2010. [Online]. Available: http://site.ebrary.com/lib/academiccompletetitles/home.action

[121] A. Mikusheva, "Limit theorems, OLS, and HAC," 2020. [Online]. Available: https://ocw.mit.edu/courses/economics/14-384-time-series-analysis-fall-2013/lecture-notes/

[122] M. Ortmanns and F. Gerfers, *Continuous-Time Sigma-Delta A/D Conversion: Fundamentals, Performance Limits and Robust Implementations*, ser. Springer Series in Advanced Microelectronics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, vol. 21. [Online]. Available: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10129683

[123] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1739–1755, 2005.

[124] M. Santina and A. R. Stubberud, "Quantization Effects," in *The Control Handbook: Control System Fundamentals*, W. S. Levine, Ed. CRC press, 2017, pp. 14–1.

[125] L. Guzzella and C. H. Onder, *Introduction to Modeling and Control of Internal Combustion Engine Systems*, 2nd ed. s.l.: Springer-Verlag, 2010. [Online]. Available: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10359800

[126] E. Bas, *Basics of Probability and Stochastic Processes.* Cham: Springer International Publishing, 2019.

[127] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: Forecasting and control*, 4th ed., ser. Wiley series in probability and statistics. Hoboken, New Jersey: Wiley a John Wiley & Sons Inc, 2008.

[128] R. G. Gallager, *Discrete Stochastic Processes*, ser. The Springer International Series in Engineering and Computer Science, Communications and Information Theory. Boston, MA: Springer, 1996, vol. 321.

[129] E. Lukacs, Z. W. Birnbaum, and E. Lukacs, *Stochastic Convergence*, 2nd ed. Burlington: Elsevier Science, 2014. [Online]. Available: http://gbv.eblib.com/patron/FullRecord.aspx?p=1901430

[130] V. Z. Marmarelis, *Nonlinear dynamic modeling of physiological systems*, ser. IEEE Press series on biomedical engineering. Hoboken, New Jersey and Piscataway, New Jersey and Piscataway, New Jersey: Wiley-Interscience IEEE Press and IEEE Xplore, 2004, vol. 14.