Contributions to numerical differentiation using orthogonal polynomials and its application to fault detection and parameter identification

Dissertation

zur Erlangung des Grades

des Doktors der Ingenieurwissenschaften

der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

und des Grades

Docteur en Automatique

der Université Paris-Saclay

im Rahmen einer Cotutelle-Vereinbarung

von

Amine Othmane

Saarbrücken

2022

Tag des Kolloquiums:	2. Dezember 2022
Dekan:	Prof. Dr. Ludger Santen
Berichterstatter:	Prof. Dr.-Ing. habil. J. Rudolph Prof. Dr. Hugues Mounier Prof. Dr. Mamadou Mboup Prof. Dr. Cédric Join
Akad. Mitarbeiter:	Dr.-Ing. Christian Bur
Vorsitz:	Prof. Dr. Kathrin Flaßkamp

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Ort, Datum

_____

Unterschrift

_____

Short summary

The reconstruction of unmeasured quantities in dynamical systems often boils down to the knowledge of derivatives of the measured system variables. This work presents a unified framework for synthesizing and analysing differentiators based on classical orthogonal polynomials. Existing approaches are extended, and their relations to established methods are investigated. Parameter selection guidelines are derived based on filter interpretations of the differentiators to achieve desired frequency-domain properties. The discussion of the discrete-time implementation emphasizes the preservation of the latter properties. A new tuning approach based on an optimization problem which requires only the measured signal is proposed. The differentiators are used for model-based fault detection problems in two experimental case studies. First, the collision of a table tennis ball with a magnetically supported plate is discussed. Then, a model-based approach for the efficient real-time detection of faults in rolling element bearings is proposed. Finally, a parameter estimation problem is discussed. This work generalises recently proposed algorithms. The derived convergence conditions are less restrictive than the previously published ones. Two experimental case studies validate the theoretical analysis. These examples underline the great potential of these methods.

Kurzfassung

Die Rekonstruktion nicht direkt gemessener Größen läuft oft auf die Kenntnis von Ableitungen gemessener Signale hinaus. In dieser Arbeit wird ein einheitlicher Rahmen für die Synthese und Analyse von auf klassischen orthogonalen Polynomen basierenden Ableitungsschätzern diskutiert. Bestehende Ansätze werden erweitert und ihre Beziehungen zu etablierten Methoden untersucht. Auf der Grundlage von Filterinterpretationen der Ableitungsschätzer werden Empfehlungen für die Parameterauswahl hergeleitet, um gewünschte Eigenschaften im Frequenzbereich zu erreichen. Bei der Diskussion der zeitdiskreten Implementierung wird auf die Bewahrung dieser Eigenschaften eingegangen. Des Weiteren wird ein neuer Parametrierungsansatz vorgeschlagen, der auf ein Optimierungsproblems basiert und nur das gemessene Signal benötigt. Die Anwendung der Ableitungsschätzer im Rahmen regelungstechnischer Aufgaben wird mittels experimenteller Fallstudien zur modellbasierten Fehlerdiagnose demonstriert. Zunächst wird die Kollision eines Tischtennisballs mit einer magnetisch gelagerten Platte diskutiert. Anschließend wird ein modellbasierter Ansatz für die effiziente Echtzeiterkennung von Fehlern in Wälzlagern vorgeschlagen. Schließlich werden Ansätze zur Parameterschätzung diskutiert. Diese Arbeit verallgemeinert kürzlich vorgeschlagene Algorithmen. Die hergeleiteten Konvergenzbedingungen sind weniger restriktiv als die zuvor veröffentlichten. Zwei experimentelle Fallstudien dienen der Validierung.

Summary

The reconstruction of unmeasured quantities in dynamical systems often boils down to the knowledge of derivatives of the measured system variables. The approximation of these derivatives in the presence of measurement disturbances is known to be challenging. However, numerical differentiation algorithms based on orthogonal polynomials and truncated generalised Fourier series may considerably simplify the problem. These differentiators are robust to measurement disturbances and may contribute to solving complex control engineering tasks. Critical challenges for the application of the methods are the selection of favourable parameters and their real-time implementation.

This work presents a unified framework for synthesizing and analysing differentiators based on classical orthogonal polynomials. Existing approaches are extended, and their relations to established methods are investigated. Differentiators based on Jacobi polynomials, also known as algebraic differentiators, form a particular class of the considered algorithms. Parameter selection guidelines are derived based on filter interpretations of the differentiators to achieve desired frequency-domain properties. The discussion of the discrete-time implementation emphasizes the preservation of the latter properties. A new tuning approach based on an optimization problem which requires only the measured signal is proposed. An experimental case study compares the performance of the differentiators in the presence of measurement disturbances. The approximation results, the computational burden, and the storage requirements are discussed in detail. Especially the latter two properties are crucial for real-time applications.

The differentiators are used for model-based fault detection problems in two experimental case studies. First, the collision of a table tennis ball with a magnetically supported plate is discussed. Only the measurement of the plate position and the applied forces are known. The proposed approach significantly reduces the computational burden and memory requirements when compared to previously considered methods. Besides, the new approach decreases the minimum detectable falling height of the ball. Then, a model-based approach for the efficient real-time detection of faults in rolling element bearings is proposed. The approach is validated using experimental data from different test benches.

Finally, a parameter estimation problem is discussed. This work generalises recently proposed algorithms. The derived convergence conditions are less restrictive than the previously published ones. Besides, this approach allows identifying a subset of parameters even if some are not excited. Two experimental case studies validate the theoretical analysis. The results are compared to those achieved using standard gradient estimators and algebraic parameter estimation methods. These examples underline the great potential of these methods.

Zusammenfassung

Die Rekonstruktion nicht direkt gemessener Größen in dynamischen Systemen läuft oft auf die Kenntnis von Ableitungen beliebiger endlicher Ordnung der gemessenen Systemvariablen hinaus. Die Approximation dieser Ableitungen bei Vorhandensein von Messstörungen ist bekanntermaßen eine Herausforderung. Numerische Ableitungsalgorithmen, die auf orthogonalen Polynomen und abgeschnittenen verallgemeinerten Fourier-Reihen basieren, können das Problem jedoch erheblich vereinfachen. Diese Ableitungsschätzer sind robust gegenüber Messstörungen und können zur Lösung komplexer regelungstechnischer Aufgaben beitragen. Kritische Herausforderungen bei der Anwendung der Methoden sind die Auswahl günstiger Parameter und ihre Implementierung in Echtzeit.

In dieser Arbeit wird ein einheitlicher Rahmen für die Synthese und Analyse von auf klassischen orthogonalen Polynomen basierenden Ableitungsschätzern diskutiert. Bestehende Ansätze werden erweitert und ihre Beziehungen zu etablierten Methoden untersucht. Ableitungsschätzer, die auf Jacobi-Polynomen basieren, auch algebraische Ableitungsschätzer genannt, bilden eine besondere Klasse der betrachteten Algorithmen. Auf der Grundlage von Filterinterpretationen der Ableitungsschätzer werden Empfehlungen für die Parameterauswahl hergeleitet, um gewünschte Eigenschaften im Frequenzbereich zu erreichen. Bei der Diskussion der zeitdiskreten Implementierung wird auf die Bewahrung dieser Eigenschaften detailliert eingegangen. Des Weiteren wird ein neuer Parametrierungsansatz vorgeschlagen, der auf ein Optimierungsproblems basiert und nur das gemessene Signal benötigt. Eine experimentelle Fallstudie vergleicht die Ergebnisse der Ableitungsschätzer in Gegenwart von gestörten Messungen. Die Approximationsergebnisse, der Rechenaufwand und die Speicheranforderungen werden ausführlich diskutiert. Insbesondere die beiden letztgenannten Eigenschaften sind für Echtzeitanwendungen entscheidend.

Die Anwendung der Ableitungsschätzer im Rahmen regelungstechnischer Aufgaben wird mittels zweier experimenteller Fallstudien zur modellbasierten Fehlerdiagnose demonstriert. Zunächst wird die Kollision eines Tischtennisballs mit einer magnetisch gelagerten Platte diskutiert. Dabei sind nur die Messung der Plattenposition und die eingeprägten Kräfte bekannt. Der vorgeschlagene Ansatz reduziert den Rechenaufwand und den Speicherbedarf im Vergleich zu den bisher betrachteten Methoden erheblich. Außerdem verringert der neue Ansatz die minimal detektierbare Fallhöhe des Balls. Anschließend wird eine modellbasiertere Methode für die effiziente Echtzeiterkennung von Fehlern in Wälzlagern vorgeschlagen. Der Ansatz wird anhand experimenteller Daten von verschiedenen Prüfständen validiert.

Schließlich werden Ansätze zur Parameterschätzung diskutiert. Diese Arbeit verallgemeinert kürzlich vorgeschlagene Algorithmen. Die hergeleiteten Konvergenzbedingungen sind weniger restriktiv als die zuvor veröffentlichten. Außerdem ermöglicht dieser Ansatz die Identifikation einer Teilmenge von Parametern, selbst wenn einige nicht angeregt werden. Zwei experimentelle Fallstudien dienen der Validierung. Die Ergebnisse werden mit mittels etablierten Gradientenschätzern und algebraischen Methoden zur Parameterschätzung erzielten Ergebnissen verglichen. Diese Beispiele unterstreichen das große Potential dieser Methoden.

Résumé

La reconstruction de grandeurs non mesurées de systèmes dynamiques se résume souvent à la connaissance de dérivées d'ordres arbitraires mais fini des variables connues du système. L'approximation de ces dérivées à partir de mesures perturbées est néanmoins connu pour être un problème complexe. Toutefois, des algorithmes de différentiation numérique basés sur des polynômes orthogonaux et des séries de Fourier généralisées tronquées peuvent le simplifier considérablement. Ces dérivateurs sont en effet robustes aux bruits de mesure et peuvent contribuer à la résolution de problèmes complexes dans le domaine de l'automatique. Néanmoins, la sélection des paramètres des dérivateurs ainsi que leurs implémentations en temps réel constituent des défis critiques lors de l'application de ces méthodes.

Ce travail présente un cadre unifié pour la synthèse et l'analyse de dérivateurs basés sur des polynômes orthogonaux classiques. Des approches existantes sont davantage développées et leurs relations avec les méthodes établies sont analysées. Les dérivateurs basés sur les polynômes de Jacobi, également appelés dérivateurs algébriques, constituent une classe particulière des algorithmes étudiés dans ce travail. Des directives pour la sélection des paramètres sont déduites à partir des interprétations des filtres des dérivateurs afin d'obtenir les propriétés souhaitées dans le domaine fréquentiel. Une nouvelle approche de paramétrage basée sur un problème d'optimisation et nécessitant uniquement le signal mesuré est proposée. Lors de la discussion de l'implémentation en temps discret, la préservation des propriétés dans le domaine fréquentiel est abordée en détail. Une étude expérimentale compare les résultats issus des différents dérivateurs. Les approximations obtenues, la charge de calcul et les besoins de mémoire sont analysés et interprétés. Les deux dernières caractéristiques sont essentielles pour les applications en temps réel.

L'application des dérivateurs dans le domaine de l'automatique est démontrée à l'aide de deux études expérimentales pour la détection de défauts avec des algorithmes basés sur des modèles physiques. Tout d'abord, la collision d'une balle de ping-pong avec une plaque à lévitation magnétique est discutée. Seules la mesure de la position de la plaque et les forces appliquées sont disponibles. L'approche proposée réduit considérablement le temps de calcul et les besoins en mémoire comparée aux méthodes considérées jusqu'à présent. En outre, la nouvelle approche réduit la hauteur de chute minimale détectable de la balle. Ensuite, une approche est proposée pour une détection efficace et en temps réel des défauts dans des roulements. Cette méthode est validée à l'aide de données expérimentales provenant de différents bancs d'essai.

Des approches d'estimation de paramètres sont ensuite discutées. Ce travail généralise des algorithmes récemment proposés. Les conditions de convergence dérivées sont moins restrictives que celles publiées précédemment. De plus, cette approche permet d'identifier un sous-ensemble de paramètres, même si certains ne sont pas excités. Deux études expérimentales valident l'analyse théorique. Les résultats sont comparés à ceux obtenus avec des estimateurs établis et des méthodes algébriques d'estimation des paramètres. Ces exemples confirment les potentiels de ces méthodes développées dans ce travail.

Contents

1 Introduction **1**

1.1 Numerical differentiation 1
1.2 Fault diagnosis 3
1.3 Parameter estimation 4
1.4 Contributions and outline of the thesis 5

2 Numerical differentiation using orthogonal polynomials **9**

2.1 The approximation problem of arbitrary functions 10
2.1.1 Sets of orthogonal functions 11
2.1.2 The approximation problem 12
2.1.3 Orthogonal polynomial sets 14
2.2 Causal real-time signal approximation 18
2.2.1 Causal approximation 18
2.2.2 Degree of exactness and delay 21
2.3 The approximation problem of derivatives of arbitrary orders 23
2.3.1 An approximation-theoretic point of view 23
2.3.2 An algebraic point of view 25
2.3.3 A system-theoretic point of view 27
2.3.4 An inverse-problem point of view 28
2.3.5 A wavelet-decomposition point of view 29
2.4 Numerical differentiation using classical orthogonal polynomials 29
2.4.1 Hermite differentiators 30
2.4.2 Laguerre differentiators 34
2.4.3 Jacobi differentiators 43
2.4.4 Bibliographical and historical comments 50
2.5 Summary and concluding remarks 51

3 Tuning and real-time implementation of the differentiators **53**

3.1 Systematic tuning of classical orthogonal differentiators 54
3.1.1 Tuning of Hermite differentiators 54
3.1.2 Tuning of Laguerre differentiators 55
3.1.3 Tuning of Jacobi differentiators 56
3.2 Implementation of classical orthogonal differentiators 59

3.2.1 Implementation of Hermite differentiators 60
3.2.2 Implementation of Laguerre differentiators 64
3.2.3 Implementation of Jacobi differentiators 72
3.3 Application of systematic tuning guidelines 73
3.3.1 Derivative approximation 74
3.3.2 Approximate inversion of analogue anti-aliasing filters 82
3.4 Automatic tuning of orthogonal differentiators 84
3.4.1 Proposed approach 85
3.4.2 Examples for automatic tuning 87
3.4.3 Concluding remarks 90
3.5 Summary and concluding remarks 92

4 Model-based detection of impulsive forces 93
4.1 Model-based collision detection on a magnetically supported plate 95
4.1.1 Collision detection and collision time estimation 96
4.1.2 Parametrisation of differentiators 98
4.1.3 Implementation of detection algorithms 101
4.1.4 Concluding remarks 103
4.2 Model-based fault detection for rolling element bearings 104
4.2.1 Modelling 105
4.2.2 Parameter estimation 106
4.2.3 Fault detection 107
4.2.4 Experimental validation 109
4.2.5 Concluding remarks 114
4.3 Summary and concluding remarks 115

5 On the problem of parameter estimation 116
5.1 Background material on Dynamic Regressor Extension and Mixing 117
5.1.1 Gradient-descent estimator and persistent excitation 118
5.1.2 Dynamic Regressor Extension and Mixing 118
5.2 Two modifications of the DREM approach 121
5.2.1 A first modification of the DREM approach 122
5.2.2 A second modification of the DREM approach 124
5.3 Validation of the approaches 126
5.3.1 Simulation study 127
5.3.2 Experimental case study 1 128
5.3.3 Experimental case study 2 132
5.4 Concluding remarks 135

6 Concluding remarks and outlook 138

A Special functions and their properties 141
A.1 Classical orthogonal polynomials 141
A.1.1 Hermite Polynomials 141
A.1.2 Laguerre polynomials 142
A.1.3 Jacobi polynomials 143

A.2 Other functions . . . 143
A.2.1 Gamma function and related functions . . . 143
A.2.2 Error function . . . 143
A.2.3 Beta function and related functions: . . . 144
A.2.4 Sinc function . . . 144
A.2.5 Bessel function of the first kind . . . 144

B The Fourier transform and its properties **145**
B.1 Properties of the Fourier transform . . . 145
B.2 Fourier transform of special functions . . . 146
B.2.1 Hermite polynomials . . . 146
B.2.2 Laguerre polynomials . . . 146

References **147**

List of acronyms

OP	orthogonal polynomial
DoE	degree of exactness
orth. diff.	orthogonal differentiator
Hermite diff.	Hermite differentiator
Lag. diff.	Laguerre differentiator
diff.	differentiator
Jac. diff.	Jacobi differentiator
HG	high-gain
BT	bilinear transform
FD	forward difference
FIR	finite-impulse response
IIR	infinite-impulse response
LTI	linear time-invariant
SNRout	output signal to noise ratio
SBR	signal to bias ratio
SNR	signal to noise ratio
CDF	cumulative distribution function
DREM	Dynamic Regressor Extension and Mixing
gDREM	generalised DREM
LRE	linear regression equation
PE	persistently excited
BIBO	bounded-input bounded-output

1

Introduction

Describing the real world in mathematical terms, i.e., deriving mathematical models from observations, has been a key to understanding natural phenomena and technological systems. Scientists and engineers employed these models to design the technologies that shaped the course of history and characterise our "modern world". Mathematical models are used, for example, in weather forecasting, to predict planet movements, analyse biomedical images, design electronic circuits, send satellites to their orbits, minimise energy consumption of machines, and detect system malfunctions.

Mathematical models of technical systems are undoubtedly crucial for most modern control methods (see, e.g., Aström and Eykhoff (1971); Isermann (2013a); Ljung (1999)). However, a wide variety of information that is required to infer, study, and use these models is not directly accessible to measurements (consider, e.g., numerical values of system parameters and system states). The missing information needs to be extracted from measured signals that are corrupted by disturbances.

It is well known that for linear and nonlinear systems, the outputs and a finite set of their derivatives play an essential role in determining unknown system quantities/variables. According to Diop and Fliess (1991a, 1991b), for instance, a nonlinear input-output system is observable if, and only if, any system variable is a differential function of the inputs and the outputs. This condition means that the system variables are functions of the inputs and the outputs and their respective derivatives of some finite order. In fact, any unknown quantity such as system parameters, faults, or disturbances, may be determined if, and only if, it is expressible as a differential function of the input and output variables (see, e.g., Fliess and Join (2003); Fliess, Join, and Sira-Ramírez (2008); Fliess and Sira-Ramírez (2003a)). Thus, the problem of extracting the required information boils down to numerical differentiation, i.e., the approximation of derivatives of noisy signals.

1.1 Numerical differentiation

The numerical estimation of derivatives of measured signals is a fundamental problem in many fields of applied mathematics and engineering. Application areas range from chemistry (see, e.g., Kiss, Lv, and Hudson (2005)), biology (see, e.g., D'Amico and Ferrigno (1992); Schäfer, Rosenblum, Abel, and Kurths (1999)), physics (see, e.g., Ahnert and Abel (2007); Raffel, Willert, and Kompenhans (2007)), and applied mathematics (see, e.g., Anderssen and Bloomfield (1974); Anderssen and de Hoog (1984); Anderssen and Hegland (1999); Cullum

(1971); Hanke and Scherzer (2001); Kantz and Schreiber (2004)) to a variety of problems in engineering (see, e.g., Diop, Grizzle, Moraal, and Stefanopoulou (1994); Levant (2003); Mboup, Join, and Fliess (2007); Mboup et al. (2020); Reger, Sira-Ramírez, and Fliess (2005); Wang, Zheng, Efimov, and Perruquetti (2018)). Estimating derivatives is, however, a challenging and ill-posed problem in the sense that a small error in the measured signal can yield notable errors in the approximated derivatives.

This problem has attracted a lot of attention over decades and different approaches have been developed. Mollification and Tikhonov regularisation methods have been derived for instance in Hao (1994); Murio (2011) and Cullum (1971); Hanke and Scherzer (2001), respectively. The discrete Savitzky-Golay filters have been developed in analytical chemistry and spectroscopy as described in Madden (1978); Savitzky and Golay (1964). Frequency-domain digital filter design techniques have been used in Chen and Lee (1995); Rader and Jackson (2006) to solve the problem. Differentiation by integration approaches, i.e., methods where the approximated derivative is computed as the integral of the measured signal multiplied with a special function, have been discussed in Cioranescu (1938); Lanczos (1956). Approaches based on the properties of orthogonal polynomials (OPs) have been discussed in Blackman (1965); Bromba and Ziegler (1983); Burch, Fishback, and Gordon (2005); Mai and Hillermeier (2008); Persson and Strang (2003); Rangarajan and Purushothaman (2005). The history of approximation formulas for higher-order derivatives by integrals involving OPs has been discussed in Diekema and Koornwinder (2012). The control community has developed approaches based on observers as in Chitour (2002); Dabroom and Khalil (1997, 1999); Levant (1998, 2003) and differential-algebraic methods, which are discussed in detail in the following.

A family of numerical differentiators (diffs.) has been derived using differential-algebraic manipulations of truncated Taylor series in Mboup et al. (2007); Mboup, Join, and Fliess (2009). These works improved the algebraic techniques developed to solve various problems in signal and image processing in Fliess, Join, Mboup, and Sira-Ramírez (2004); Fliess and Sira-Ramírez (2003b, 2004). A least squares interpretation can be attached to the obtained derivative estimates as shown in Mboup (2009); Mboup et al. (2007); Mboup et al. (2009); Nöthen (2007). Time-domain derivations of these diffs., called in the sequel algebraic diffs., can be found in Kiltz (2017); Liu, Gibaru, and Perruquetti (2011a). The design process of these diffs. involves five parameters, the judicious choice of which is crucial to achieve a good accuracy of the estimates. The tuning of the diffs. has been discussed in Kiltz (2017); Kiltz and Rudolph (2013); Liu et al. (2011a); Mboup et al. (2009); Mboup and Riachy (2014, 2018); Othmane, Mounier, and Rudolph (2021). In particular, it has been shown in Kiltz (2017); Kiltz and Rudolph (2013) that the diffs. can be approximated as low-pass filters of an arbitrary order, and the parameters can be computed from a desired cutoff frequency and the desired filter order. A summary of the tuning guidelines and their application on example case studies can be found in Othmane, Kiltz, and Rudolph (2022). An exhaustive list of applications including parameter estimation, state reconstruction, feedback control, fault diagnosis, anomaly detection, fault-tolerant control, and model-free control has been provided in Othmane et al. (2022). A MATLAB and Python toolbox implementing all necessary functions for the design, analysis, and discretisation of the filters has been released with the discussions in Othmane et al. (2022) and is accessible in Othmane (2021).

The relations to existing methods such as deadbeat state estimation, Savitzky-Golay filters, Legendre filters, and mollification have been discussed in Kiltz (2017); Othmane et al. (2022);

Reger and Jouffroy (2009); Riachy, Efimov, and Mboup (2016). A detailed discussion of the historical developments of these diffs. with an analysis and summary of the relations to established estimation approaches can be found in Othmane et al. (2022). It has been shown that these diffs. outperform high-gain and sliding mode diffs. in Liu (2011); Othmane, Rudolph, and Mounier (2021a); Sidhom (2011), for example. The result of algebraic diffs. is comparable to that of a high-gain differentiator when a delay-free estimation is considered (see Othmane, Rudolph, and Mounier (2021a)). However, when a small but known delay is tolerated, the estimation result is greatly improved. It has also been claimed that with the systematic tuning guidelines, the parametrisation of algebraic diffs. is simpler than that of high-gain or sliding mode diffs. In the experimental case study Yan, Primot, and Plestan (2014), algebraic diffs. have been compared to sliding mode ones for the estimation of velocities and accelerations in pneumatic systems. It has been argued that for low frequency signals, the first approach outperforms the second and it is the opposite for high frequencies. However, this observation is probably due to the large filter window, which corresponds to a low cutoff frequency. Moreover, the observed deterioration of the results for the algebraic approach with reduced sampling frequency can be overcome by a suitable and systematic parametrisation of the diffs. as will be shown in this work.

A different class of algebraic diffs. has been proposed in Ushirobira (2018); Ushirobira and Quadrat (2016). Therein, a series expansion of the signals in sums of Laguerre and Hermite polynomials have been considered. Then, differential-algebraic manipulations are used in the operational domain to design estimators for the derivative of interest. However, neither an analysis of the estimation error and delay nor the tuning and implementation of the diffs. are discussed. These missing investigations are presented in this thesis.

1.2 Fault diagnosis

The growing complexity of technical systems and the implementation of modern control algorithms allow companies to increase their process efficiency, profitability, and product quality. However, this complexity increases the potential fault sources and may lead to severe damage in safety-critical applications, for example. Here, a fault should be understood as any malfunction deteriorating the performance of systems. Frank (1990) distinguishes three types of faults: actuator faults (malfunctions of the actuators), component faults (malfunctions in the other components affecting the dynamics of the process), and instrument faults (malfunctions of the sensors). Faults need to be detected, i.e., the occurrence of faults needs to be determined, then isolated, i.e., the kind and the location of the fault need to be determined, and finally identified, i.e., the size and time behaviour need to be determined. These three tasks can be summarised as the process of fault diagnosis (see, e.g., Isermann and Ballé (1997)). Fast and accurate fault diagnosis algorithms are crucial to increase plant reliability and reduce the risk of safety hazards.

Numerous approaches have been developed for fault diagnosis over the last decades: Parity space methods (see, e.g., (Isermann, 2006, Ch. 10) and Gertler (1991)), observer-based methods (see, e.g., (Isermann, 2006, Ch. 11) and Frank (1990)), fault-detection filters (see, e.g., Frank (1992); Willsky (1976)). A vast body of literature is available on the topic and the main methods have been well summarised in Frank (1990); Gertler (1991); Hwang, Kim, Kim, and Seah (2010); Isermann (2006); Patton, Frank, and Clark (1989); Willsky (1976). Algebraic

fault detection methods have been initiated in Join (2002) and further discussed in Fliess and Join (2003); Fliess, Join, and Sira-Ramírez (2004). As for the algebraic derivative estimation approaches, fault diagnosis algorithms are derived using differential-algebraic manipulations in the operational domain.

Fault diagnosis methods have also been investigated for nonlinear systems (see, e.g., Alavi and Saif (2010); De Persis and Isidori (2001, 2002); Jiang, Staroswiecki, and Cocquempot (2006); Yan and Edwards (2007); Yang and Saif (1996)). These methods require a considerable implementation and comprehension effort. In contrast, methods based on numerical differentiation and especially algebraic diffs. have shown a great success in solving challenging problems mainly due to their simplicity and efficient implementations. These approaches and their implementation have been discussed in Ahmed et al. (2016); Fliess, Join, and Mounier (2004); Fliess et al. (2008); Kiltz, Join, Mboup, and Rudolph (2014); Lomakin and Deutscher (2020a, 2022); Mai and Hillermeier (2010); Mai and Hillermeier (2010); Tisserand, Lezama, Schweitzer, and Berviller (2015), for instance. An extensive list of applications in various domains is provided in Othmane et al. (2022), for example.

1.3 Parameter estimation

When working with a model to predict, explain, or control system dynamics, estimating the numerical values of model parameters is an important task. The literature on parameter estimation is extensive. For general textbooks on the subject from the control theory perspective the reader is referred to Garnier and Wang (2008); Isermann (2013a, 2013b); Ljung (1999); Unbehauen and Rao (1987), for example. Despite the long-standing history of the topic, it still attracts considerable attention and effort.

Some of the recently-proposed methods are again based on differential-algebraic manipulations and have their roots in the work Fliess and Sira-Ramírez (2003a). A least squares interpretation and a finite-impulse response (FIR) filter implementation of these algebraic methods have been discussed in Mboup (2009). Problems in various domains ranging from electrical to automotive engineering have been solved with these methods. The online estimation of a load of a boost converter has been discussed in Gensior, Weber, Rudolph, and Güldner (2008) for example. The parameters of rigid bodies have been identified in Lomakin and Deutscher (2020b). The estimation of the parameters of harmonic signals has been discussed in Liu, Gibaru, Perruquetti, Fliess, and Mboup (2008); Mboup (2009); Ushirobira, Perruquetti, and Mboup (2016) for example. A major advantage of these approaches is the simplicity of implementation.

In Aranovskiy, Bobtsov, Ortega, and Pyrkin (2016b, 2017), a new approach for parameter estimation has been proposed for models in the form of linear regression equations (LRE). The convergence of the approach does not depend on a persistence of excitation condition, which is required for the convergence of standard gradient estimators (see, e.g., (Slotine & Li, 1991, Sec. 8.7.3) or (Sastry & Bodson, 1989, Sec. 2.5)). This condition, which is ubiquitous in most adaptive control algorithms (see, e.g., Åström and Wittenmark (2013); Narendra and Annaswamy (2012); Slotine and Li (1991)), is rarely verified in applications. The new approach called Dynamic Regressor Extension and Mixing (DREM) consists of two steps. First, new regression equations are generated by using suitable operators applied on the measurement data. In a second step, these equations are combined in a way to ensure that each parameter

satisfies an independent regression equation. The parameters can then be estimated using a gradient estimator for example and the convergence depends on the non-square integrability of the determinant of a specific matrix. The approach has attracted a lot of attention (see, e.g., Belov, Ortega, and Bobtsov (2018); Ortega, Aranovskiy, Pyrkin, Astolfi, and Bobtsov (2021); Ortega, Gerasimov, Barabanov, and Nikiforov (2019); Ortega, Nikiforov, and Gerasimov (2020); Ortega, Praly, Aranovskiy, Yi, and Zhang (2018); Yi and Ortega (2022)) and has been applied to different problems (see, e.g., Aranovskiy, Bobtsov, Ortega, and Pyrkin (2016a); Bazylev, Doria-Cerezo, Pyrkin, Bobtsov, and Ortega (2017); A. Bobtsov, Bazylev, Pyrkin, Aranovskiy, and Ortega (2017); A. A. Bobtsov, Pyrkin, Ortega, and Vedyakov (2018); Yi, Ortega, Siguerdidjane, and Zhang (2018)).

1.4 Contributions and outline of the thesis

This thesis can be divided into two main parts. The first part consists of the Chapters 2 and 3 and deals with the derivations, interpretations, tuning, and implementation of numerical differentiation algorithms based on OPs. Chapters 4 and 5 form the second part and discuss different applications of these diffs. in the context of fault detection and parameter identification. In the following, the contributions are outlined and clarified.

Chapter 2: Numerical differentiation using orthogonal polynomials

This chapter presents the general framework for the design of diffs. based on orthogonal functions. In particular, diffs. based on classical OPs are derived, reviewed, and analysed in the time and the frequency domains. Different interpretations for these numerical differentiation schemes are provided.

Section 2.1 provides a basic introduction to the approximation problem of arbitrary functions by recalling the mathematical background on which the remaining sections depend. Sets of orthogonal functions and generalised Fourier series play a predominant role. The notation required for the remaining parts of the chapter is thus introduced.

Section 2.2 uses the methods from Section 2.1 to the causal real-time approximation of signals. In particular, general time transformations are introduced to map intervals of interest to the orthogonality intervals of orthogonal functions. These transformations may depend on different parameters. Their influence on the approximation errors is analysed for general orthogonal functions. The estimation delay and the degree of exactness of an approximation scheme are then introduced. It has already been shown in Mboup et al. (2007); Mboup et al. (2009) that orthogonal differentiators (orth. diffs.) based on Jacobi OPs can yield a delay in the estimation. Here, the delay is derived for general time transformations. The degree of exactness, first introduced in Kiltz (2017), describes the degree of a polynomial the approximation of which is exact up to a known delay (see also Othmane et al. (2022)).

In Section 2.3, the approximation of arbitrary derivatives of signals is considered using sets of orthogonal functions. The results of the preceding two sections are used and different interpretations for the diffs. are given. The diffs. are then considered from an algebraic point of view where the key ideas are the interpretation of undesired derivatives as disturbances and their annihilation. Then, system theoretic, inverse problem, and wavelet-decomposition viewpoints are presented.

Section 2.4 applies the methods from Sections 2.1 to 2.3 to sets of classical orthogonal polynomials. Thus, diffs. based on Hermite, Laguerre, and Jacobi polynomials are considered and called in the sequel Hermite differentiators (Hermite diffs.), Laguerre differentiators (Lag. diffs.), and Jacobi differentiators (Jac. diffs.), respectively. Their properties in the time and frequency domains are investigated. The Jac. diffs. have already been derived in Mboup et al. (2007); Mboup et al. (2009) and analysed in Kiltz (2017); Kiltz and Rudolph (2013); Liu et al. (2011a); Mboup and Riachy (2014, 2018) and their properties are only recalled here. Contrarily, the proposed Lag. diffs. and Hermite diffs. are more general than those from Ushirobira (2018) in the sense that different tuning parameters are considered. Their influence on the approximation errors, delay, and degree of exactness will also be analysed. In Ushirobira (2018); Ushirobira and Quadrat (2016), the signal is first expanded into a series of Laguerre or Hermite polynomials. Then, annihilators are derived to identify the coefficients of the required derivatives. Section 2.4 considers directly the Taylor series expansion for the derivation of the annihilators required for the Lag. diffs. as in the works Mboup et al. (2007); Mboup et al. (2009) on Jac. diffs.

Chapter 3: Tuning and real-time implementation of classical orthogonal differentiators

The diffs. developed in Chapter 2 depend on numerous parameters that affect the approximation error and delay. Finding good parameter compromises to achieve a satisfying accuracy of the approximated derivatives without relying on trial-and-error approaches is crucial. Besides, in most applications, the measured signals are available at discrete sampling instants only. The discretised diffs. have to preserve the properties of the continuous-time ones. Achieving this can be challenging, especially for low sampling rates. All these issues are addressed in detail in Chapter 3.

Section 3.1 analyses the systematic tuning of diffs. based on OPs. It is shown that these diffs. can be interpreted as low-pass filters, and the parameters can be computed from a desired cutoff frequency. An arbitrary filter order, i.e., stopband slope, can be assigned to the Jac. diffs. and Lag. diffs. These considerations greatly simplify the tuning process as already discussed for Jac. diffs. in Kiltz (2017); Kiltz and Rudolph (2013); Othmane et al. (2022).

The discrete-time implementation of the diffs. is analysed in Section 3.2. Different discretisation schemes are considered and their effects, especially on the preservation of frequency domain properties, are investigated. It is shown that all considered diffs. can be implemented as FIR filters or sufficiently closely approximated by the latter. An error norm, initially introduced in Kiltz (2017) for the truncation of the filter window length of Jac. diffs., is used here to approximate Lag. diff. and Hermite diffs. by FIR filters. The error norm will be important when experimental approximation results are compared. Besides, it is illustrated that for a particular parameter choice, Lag. diffs. can be efficiently approximated by dynamic systems, the solutions of which converge exponentially to the approximation of the former. This result enables the analysis of the relations to established approaches such as state-variable filters and high-gain (HG) diffs. In particular, it is shown that the latter are a special case of Lag. diffs., which allows the quantification of their estimation delay. Simple academic examples are used for illustration.

Section 3.3 applies the previously mentioned tuning and discretisation guidelines to two experimental case studies. First, the derivatives of a disturbed signal are approximated. Then,

the approximate inversion of analogue anti-aliasing filters is considered as presented in Kiltz et al. (2019); Othmane et al. (2022). The approximation results are compared qualitatively and quantitatively for the different diffs. The computational burden and storage requirements for each approach is analysed in detail. These considerations are especially important for industrial applications.

An entirely new and different tuning approach is proposed in Section 3.4. Labelled as "automatic tuning approach", this method is based on an optimisation problem requiring only the measured signal as input and yields the filter parameters as output. An experimental case study shows that the approach furnishes excellent results.

Chapter 4: Model-based detection of impulsive forces

While Chapter 3 illustrates the systematic tuning guidelines in detail on some relatively simple examples, the current chapter applies the methods to two advanced experimental problems. The results validate the theoretical analysis and support the importance of these numerical differentiation schemes for practical problems.

Section 4.1 considers the collision detection of a table tennis ball weighing $2.7\,\mathrm{g}$ with a magnetically levitated plate with a mass of $4.7\,\mathrm{kg}$. A detailed description of the experimental setup can be found in Kiltz et al. (2014); Kiltz and Rudolph (2013). The precise estimation of the collision instant is also investigated. The problem is solved by using the measurement of the vertical position of the plate only. The approach developed in Kiltz (2017); Kiltz and Rudolph (2013) is compared to a new one based on Lag. diffs. and notch filters. This new approach requires $95.33\,\%$ fewer arithmetic operations than the approach from Kiltz and Rudolph (2013) and $60.77\,\%$ less than Kiltz (2017). The number of filter coefficients to be stored is reduced by $96.14\,\%$ compared to that in Kiltz and Rudolph (2013) and by $75.58\,\%$ compared to that in Kiltz (2017). Besides, the new approach decreases the minimum detectable falling height of the ball by $11.7\,\%$.

In Section 4.1, a model-based approach for the efficient real-time detection of faults in rolling element bearings is presented. The algorithm is based on a model for the vibrations and efficient numerical differentiation schemes using Jac. diffs. and has been first published in Othmane and Rudolph (2021). The method includes a failure detection threshold specifying the probability of false alarms. The approach is validated by using experimental data[1] from different research centres (see CWRU (2015); Daga, Fasana, Marchesiello, and Garibaldi (2009, 2019); Nectoux et al. (2012); Smith and Randall (2015)).

Chapter 5: On the problem of parameter estimation

The parameter estimation problem is considered in Chapter 5, which consists of three parts. Especially, systems linear in their parameters and LRE are discussed in detail within the framework of the DREM methodology. Section 5.1 recalls some background material on DREM, bibliographical comments, and remarks on using orth. diffs. within this framework.

Two modifications of the DREM parameter estimation approach are proposed in Section 5.2. In the first one, an alternative derivation of the parameter estimation approach

[1]The experimental validation of the approach for the fault detection in rolling element bearing elements developed in this thesis would not have been possible without the use of the publicly available experimental data sets. The author of this thesis is greatly indebted to the research centres providing them.

is provided, which yields weaker convergence conditions than the one by Aranovskiy et al. (2016b, 2017); Ortega et al. (2021). The previously published approaches in the latter references are shown to be a special case of the new one. In the second modification, QR decompositions with column pivoting are used to split the system parameters into excited and non-excited ones. This adaptation allows identifying a subset of parameters even if others are not excited. The advantages of the approaches are discussed in simple examples. A counter-example is provided, where none of the DREM approaches yield a convergence of the estimates. Contrarily, an algebraic method (see, e.g., Mboup et al. (2009)) yields good results.

Section 5.3 validates the new techniques. First, a simulation study is provided to show the properties of the second modification. Then, two experimental case studies investigate the use of Jac. diff. as example orth. diffs. with the proposed modifications. The first one examines the estimation of frequencies from a disturbed measurement of a linear combination of two harmonic functions. The approach requires estimates of the derivatives up to the fourth order. The low Nyquist frequency equal to $100\,\mathrm{Hz}$ in comparison to the highest frequency equal to $80\,\mathrm{Hz}$ in the signal makes the design of diffs. challenging. The method is compared to the algebraic identification procedure developed in Mboup (2009). In the second case study, the parameters of an electro-mechanical system are identified. The estimates are compared to those from a standard gradient estimator. For validation, recorded input trajectories are used to compare simulation results to measurements. These examples confirm the excellent results that can be achieved using the new methods.

Chapter 6: Conclusion

Chapter 6 summarises the contributions of this thesis and provides open problems and further ideas for future research.

Appendices

Appendices A and B contain supplementary material regarding special functions and the Fourier transform.

2

Numerical differentiation using orthogonal polynomials

The approximation of functions using sets of basis functions has been studied for centuries due to its importance in solving numerous applications in engineering and physics (see, e.g., Davis and Hersh (1981)). This chapter first recalls the mathematical background, specifically generalised Fourier series and orthogonal sets, required to derive function approximations based on least squares optimisation problems.

The approximation methods are then used to derive approaches for the causal approximation of arbitrary derivatives of signals affected by additive disturbances. The main ideas here are based on time transformations mapping some orthogonality intervals of functions to an interval containing a known history of the signals of interest. These transformations depend on parameters that shift and dilate or shift and contract the intervals. The effects of these parameters are investigated. In the detailed analysis of the approximation errors, an intriguing yet interesting result will be discussed: Admitting a small but precisely known delay in the causal approximation of a signal or any of its derivatives reduces the order of the approximation error and significantly increases the robustness in the presence of additive disturbances. This has been first observed for example in Mboup et al. (2007); Mboup et al. (2009) for differentiators (diffs.) based on Jacobi orthogonal polynomials (OPs).

The approximation methods are subsequently applied to sets of classical orthogonal polynomials that include Hermite, Laguerre, and Jacobi polynomials, which have as special case the Gegenbauer, Chebyshev, and Legendre polynomials (see, e.g., Szegö (1939) for a detailed discussing on these polynomials). It is shown that the resulting approximation methods are linear filters. The kernel of theses filters act on the sought derivatives whereas the derivatives of the kernels act on the measured signals. The properties of these filters in the time and the frequency domains are discussed in detail. Moreover, it is shown that the considered diffs. can be derived using differential-algebraic manipulations of truncated Taylor series in the operational domain. These ideas root[1] in the seminal works and ideas of Fliess and his co-authors on parameter estimation in Fliess and Sira-Ramírez (2003a).

While the diffs. based on Jacobi polynomials have initially been derived in Mboup et al. (2007); Mboup et al. (2009) and investigated in Kiltz (2017); Kiltz and Rudolph (2013); Liu (2011); Liu, Gibaru, and Perruquetti (2010); Liu et al. (2011a); Liu, Gibaru, and Perruquetti (2011b); Mboup (2009), those based on Laguerre and Hermite polynomials discussed here are more general than the diffs. derived in Ushirobira (2018); Ushirobira and Quadrat (2016).

[1] See the survey Othmane et al. (2022) for a detailed discussion of the history of diffs. based on Jacobi diffs.

Their properties are studied in great detail. A detailed discussion of the historical development of the diffs. based on Jacobi polynomials and their application can be found in the survey Othmane et al. (2022) and only a summary is presented here.

Chapter content

2.1 The approximation problem of arbitrary functions 10
2.1.1 Sets of orthogonal functions 11
2.1.2 The approximation problem 12
2.1.3 Orthogonal polynomial sets 14
2.2 Causal real-time signal approximation 18
2.2.1 Causal approximation 18
2.2.2 Degree of exactness and delay 21
2.3 The approximation problem of derivatives of arbitrary orders 23
2.3.1 An approximation-theoretic point of view 23
2.3.2 An algebraic point of view 25
2.3.3 A system-theoretic point of view 27
2.3.4 An inverse-problem point of view 28
2.3.5 A wavelet-decomposition point of view 29
2.4 Numerical differentiation using classical orthogonal polynomials 29
2.4.1 Hermite differentiators 30
2.4.2 Laguerre differentiators 34
2.4.3 Jacobi differentiators 43
2.4.4 Bibliographical and historical comments 50
2.5 Summary and concluding remarks 51

2.1 The approximation problem of arbitrary functions

Consider a function $f : \mathcal{I}_1 \to \mathcal{I}_2$, with $\mathcal{I}_1, \mathcal{I}_2 \subseteq \mathbb{R}$, satisfying some integrability conditions to be specified later. The expansion of f in a convergent[2] infinite series

$$\sum_{n=0}^{\infty} a_n \phi_n, \tag{2.1}$$

with $a_n \in \mathbb{R}$ and $\phi_n : \mathcal{I}_1 \to \mathcal{I}_2$ to be specified, has been discussed for centuries. Preliminary results can be found in the works of Euler and Bernoulli on vibrating strings developed in the 18-th century[3] where ϕ_n are trigonometric functions. Substantial improvements have been made by Fourier in his works Fourier (1822); Fourier and Freeman (1878) on heat conduction in the 19-th century for trigonometric functions ϕ_n as well. The concepts of Fourier have then been rigorously analysed and generalised in the 20-th century when the orthogonality property of functions has been understood more thoroughly. This has led to powerful methods for the

[2]Different convergence definitions (e.g., convergence in the mean, uniform convergence, absolute convergence, pointwise convergence) can be considered for an infinite series of the form (2.1). A series might convergence if one definition is used and diverge for an other. This will be addressed in detail in the sequel.

[3]The reader is referred to the beautiful work Davis and Hersh (1981) where the history of Fourier expansion has been presented, and the evolution of the mathematical thinking has been presented.

approximation of arbitrary functions. Some of these approaches rely on polynomial functions ϕ_n known as OPs, with very interesting properties.

The relevant theoretical background on function approximation and infinite series of the form (2.1) used in the current work is briefly reviewed in the following subsections. For a detailed discussion of function approximation, orthogonal systems, and general OPs the reader is referred to the standard works Bateman and Erdélyi (1953); Courant and Hilbert (1924); Kaczmarz and Steinhaus (1951); Lebedev and Silverman (1965); Szegö (1939); Tricomi (1970).

2.1.1 Sets of orthogonal functions

Let w be a non-negative function defined on an interval $\mathcal{I}_w = (a, b) \subseteq \mathbb{R}$ and

$$f_1, f_2 : \mathcal{I}_w \to \mathbb{R}$$

two real functions for which $w^{\frac{1}{2}} f_1$ and $w^{\frac{1}{2}} f_2$ are square integrable, i.e., the Lebesgue integrals

$$\int_a^b \left(w^{\frac{1}{2}}(\tau) f_1(\tau) \right)^2 \mathrm{d}\tau \quad \text{and} \quad \int_a^b \left(w^{\frac{1}{2}}(\tau) f_2(\tau) \right)^2 \mathrm{d}\tau$$

exist. Then, the scalar product of the functions f_1 and f_2 denoted by $\langle f_1, f_2 \rangle_w$ is defined by the Lebesgue integral[4]

$$\langle f_1, f_2 \rangle_w = \int_a^b w(\tau) f_1(\tau) f_2(\tau) \mathrm{d}\tau. \tag{2.2}$$

This scalar product induces the norm and the distance

$$\|f_1\|_w = \sqrt{\langle f_1, f_1 \rangle_w}, \tag{2.3}$$

$$\mathrm{d}_w(f_1, f_2) = \|f_1 - f_2\|_w, \tag{2.4}$$

respectively. The function w is called the weight function of the interval $\mathcal{I}_w$, or shortly weight function.

Denote by $\mathcal{L}^2_w(\mathcal{I}_w)$ the class of all functions $f : \mathcal{I}_w \to \mathbb{R}$ for which the integral $\int_a^b w(\tau) f^2(\tau) \mathrm{d}\tau$ exists. Two distinct functions $f_1, f_2 \in \mathcal{L}^2_w(\mathcal{I}_w)$ are said to be orthogonal, if their scalar product vanishes, i.e., $\langle f_1, f_2 \rangle_w = 0$. Orthonormal sets can now be defined using these concepts.

Definition 2.1: Orthonormal set (Szegö (1939))

An orthonormal set of functions $\Phi_w = \{\phi_n\}_{n=0}^{l}$, l finite or infinite, associated with a weight function w is defined by

$$\langle \phi_n, \phi_m \rangle_w = \int_a^b w(\tau) \phi_n(\tau) \phi_m(\tau) \mathrm{d}\tau = \delta_{nm}, \quad n, m = 0, 1, \ldots, l,$$

where $\delta_{nm} = 0$ or 1, according as $n \neq m$, or $n = m$. Here ϕ_n is real-valued and belongs to the class $\mathcal{L}^2_w(\mathcal{I}_w)$.

[4] When f_1 and f_2 are complex functions, f_2 must be replaced in (2.2) by its complex conjugate. The reader is referred to Szegö (1939) for a more general definition of the scalar product by a Stieltjes-Lebesgue integral. This definition is required for the analysis of functions of discrete variables, for example.

The elements of any finite subset of an orthogonal set are linearly independent (see, e.g., (Bateman & Erdélyi, 1953, P. 154)).

The previously mentioned trigonometric functions used in the early works of Euler, Bernoulli, and Fourier form such orthonormal sets. The approximation of arbitrary functions using these sets is analysed and the concept of best approximation is introduced.

2.1.2 The approximation problem

In various engineering and scientific problems an approximation $\hat{f}_N \in \mathcal{A}_w$ of an element $f \in \mathcal{L}^2_w(\mathcal{I}_w)$ is required, with $\mathcal{A}_w$ a finite dimensional subspace of $\mathcal{L}^2_w(\mathcal{I}_w)$ and $w : \mathcal{I}_w \to \mathbb{R}$ a weight function. The accuracy of such an approximation can be evaluated using the distance $\mathrm{d}_w(f, \hat{f}_N)$ between f and $\hat{f}_N$ defined in (2.4). The approximation problem can then be formulated as follows.

Problem 2.1: The approximation problem

For a given $f \in \mathcal{L}^2_w(\mathcal{I}_w)$ determine an element $\hat{f}_N \in \mathcal{A}_w \subset \mathcal{L}^2_w(\mathcal{I}_w)$, with $\mathcal{A}_w$ finite dimensional, such that

$$\mathrm{d}_w(f, \hat{f}_N) \leqslant \mathrm{d}_w(f, \hat{h}) \tag{2.5}$$

for all $\hat{h} \in \mathcal{A}_w$. An element $\hat{f}_N$ satisfying (2.5) is called a best approximation of f with respect to $\mathcal{A}_w$.

A simple yet powerful linear approximation scheme can be derived using the orthonormal sets introduced in Section 2.1.1 to solve Problem 2.1. Let $\Phi_w = \{\phi_n\}_{n=0}^N \subset \mathcal{L}^2_w(\mathcal{I}_w)$, with N finite, be an orthonormal set and consider the linear approximation Ansatz

$$\hat{f}_N = \sum_{n=0}^{N} c_n \phi_n, \quad c_n \in \mathbb{R}.$$

Then, Problem 2.1 can be solved by minimising the square of the distance $\mathrm{d}_w(f, \hat{f}_N)$ given as

$$\left(\mathrm{d}_w(f, \hat{f}_N)\right)^2 = \int_a^b w(\tau) \left(f(\tau) - \hat{f}_N(\tau)\right)^2 \mathrm{d}\tau = \|f\|_w^2 - \sum_{n=0}^{N} f_n^2 + \sum_{n=0}^{N} (c_n - f_n)^2, \quad f_n = \langle f, \phi_n \rangle_w, \tag{2.6}$$

with respect to the coefficients c_n. This corresponds to an approximation of f in the mean. The approximation problem is solved if, and only if, the coefficients c_n satisfy

$$c_n = f_n = \int_a^b w(\tau) f(\tau) \phi_n(\tau) \mathrm{d}\tau = \langle f, \phi_n \rangle_w .$$

Thus, the function $\hat{f}_N$ minimising the square of the distance $\mathrm{d}_w(f, \hat{f}_N)$ is called in the sequel best approximation and corresponds to the N-th order truncation of the Fourier expansion defined below.

Definition 2.2: Fourier expansion (Szegö (1939))

Let $\Phi_w = \{\phi_n\}_{n=0}^{N}$, with N finite or infinite, be a given orthonormal set associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$. To an arbitrary real-valued function $f \in \mathcal{L}_w^2(\mathcal{I}_w)$ let there correspond the formal Fourier expansion[a]

$$f \sim \sum_{n=0}^{N} f_n \phi_n.$$

The coefficients f_n, called the Fourier coefficients of f with respect to Φ_w, are defined by $f_n = \langle f, \phi_n \rangle_w$.

[a] If two sequences z_n and w_n have the property that $w_n \neq 0$ and the sequence $|z_n| / |w_n|$ has finite positive limits of indetermination, the notation $z_n \sim w_n$ is used. The reader is referred to Szegö (1939) for more details.

Since the squared distance between a function f and its best approximation $\hat{f}_N$ is positive it is clear from (2.6) that

$$\sum_{n=0}^{N} f_n^2 \leqslant \|f\|_w^2 .$$

The right side of the inequality is independent of N and it can be concluded that the series $\sum_{n=0}^{\infty} f_n^2$ converges and the inequality

$$\sum_{n=0}^{\infty} f_n^2 \leqslant \|f\|_w^2 ,$$

known as Bessel's inequality (see, e.g., (Bateman & Erdélyi, 1953, P. 157)), is recovered. If the equality holds, i.e.,

$$\sum_{n=0}^{\infty} f_n^2 = \|f\|_w^2 , \tag{2.7}$$

also known as Parseval's formula, the system Φ_w is called closed in $\mathcal{L}_w^2(\mathcal{I}_w)$ and it follows that

$$\lim_{N \to \infty} \mathrm{d}_w(f, \hat{f}_N) = 0.$$

When this equality is satisfied, the partial sum $\hat{f}_N = \sum_{n=0}^{N} f_n \phi_n$ converges in the mean to f. This is discussed in detail in (Bateman & Erdélyi, 1953, Sec. 10.3), for example.

From the preceding developments it can be concluded that every $f \in \mathcal{L}_w^2(\mathcal{I}_w)$ can be approximated arbitrarily closely, in the mean, by a truncated Fourier expansion with respect to any closed orthonormal set, i.e.

$$0 \leqslant \mathrm{d}_w(f, \hat{f}_N) \leqslant \mathrm{d}_w(f, \hat{f}_{N+1}) \quad \text{and} \quad \forall \epsilon > 0 \; \exists N_0 \in \mathbb{N} \mid N > N_0 : \mathrm{d}_w(f, \hat{f}_N) < \epsilon.$$

Moreover, every closed orthonormal set Φ_w is also complete, i.e., if $\langle f, \phi \rangle_w = 0$, for all elements $\phi \in \Phi_w$, then f vanishes almost everywhere.

Remark 2.1: Convergences properties

The best approximation $\hat{f}_N$ of f is said to converge pointwise to f in $\mathcal{I}_w$, if

$$\lim_{N\to\infty} \hat{f}_N(\tau) = f(\tau)$$

for all $\tau \in \mathcal{I}_w$. This is a stronger property than the convergence in the mean and requires much more restrictive assumptions on f. This property is only addressed in Section 2.1.3 for special choices of orthonormal sets formed by classical OPs. General results can be found in (Szegö, 1939, Ch. X) and (Tricomi, 1970, Ch. V&VI), for example.

The orthonormal set Φ_w used for the series expansion of f has to be judiciously chosen and can highly influence the approximation results and the computational burden required for its online implementation. The Weierstrass approximation theorem (see, e.g., (Szegö, 1939, Sec. 1.3)) stating that every continuous function defined on a closed interval can be uniformly and arbitrarily closely approximated by a polynomial function motivates the use of polynomials for the generation of orthonormal sets.

2.1.3 Orthogonal polynomial sets

Denote by π_N and w the set of polynomials of degree N and a weight function defined on $\mathcal{I}_w = (a, b)$, respectively. An element $p_n \in \pi_N$ is a polynomial of degree $n \in \{0, \ldots, N\}$ whose leading coefficient, denoted by k_n, is not zero.

Definition 2.3: Orthogonal polynomial set

An orthogonal polynomial set $\Pi_w = \{p_n\}_{n=0}^{l}$, l finite or infinite, associated with a weight function w defined on $\mathcal{I}_w = (a, b) \subseteq \mathbb{R}$ is defined by

$$\langle p_n, p_m \rangle_w = \int_a^b w(\tau) p_n(\tau) p_m(\tau) \mathrm{d}\tau = 0 \quad \text{for} \quad n \neq m.$$

Here p_n is a polynomial of degree n and belongs to the class $\mathcal{L}_w^2(\mathcal{I}_w)$. It is called an OP associated with w on $\mathcal{I}_w$.

General properties

An orthogonal polynomial set Π_w associated with w is uniquely determined up to a non-zero scaling factor. Since every $p_n \in \Pi_w$ has the degree n, every polynomial of degree n can be represented as a linear combination of p_0, p_1, …, p_n. Thus, p_n is orthogonal to any polynomial of lower degree.

The elementary properties of the zeros of the polynomials in an OP set Π_w are summarised in the following theorems.

Theorem 2.1: Location of zeros (Szegö, 1939, Thm. 3.3.1)

The zeros of an OP p_n associated with a weight function w on the interval $\mathcal{I}_w = (a, b)$ are

real and distinct and are located in the interior of the interval $\mathcal{I}_w$.

Theorem 2.2: Separation of zeros (Szegö, 1939, Thm. 3.3.2)

Let $z_1 < z_2 < \ldots < z_n$ be the zeros of an OP p_n associated with a weight function w on the interval $\mathcal{I}_w = (a, b)$, $z_0 = a$, and $z_{n+1} = b$. Then each interval (z_m, z_{m+1}), $m \in \{0, \ldots, n\}$, contains exactly one zero of p_{n+1}.

In Section 2.1.2 it has been shown that for closed orthonormal sets Parseval's formula (2.7) holds and functions can be approximated arbitrarily closely, in the mean, by truncated Fourier expansions. It follows from (Szegö, 1939, Thm. 3.1.5) that the orthogonal polynomial set Π_w associated with w is closed for every finite interval $\mathcal{I}_w$. This is not necessarily true in general for infinite intervals, as discussed in (Szegö, 1939, Sec. 5.7). However, classical OP sets introduced in the following subsection are closed.

Definition and properties of classical orthogonal polynomials

The most widely used OPs are known as classical OPs and include Hermite, Laguerre, and Jacobi polynomials, which have as special case the Gegenbauer, Chebyshev, and Legendre polynomials. These polynomials will first be defined and then used for the derivation of causal signal estimators. A detailed investigation of the properties of these polynomials can be found in Szegö (1939), for example.

Let Π_w be a set of classical OPs associated with a weight function w defined on an interval $\mathcal{I}_w = (a, b)$. Let $p_n \in \Pi_w$ and denote by h_n the norm of p_n defined in (2.3), i.e., $h_n = \|p_n\|_w$. The weight functions, orthogonality intervals, and notations of the classical polynomials are summarised[5] in Table 2.1. All these sets formed by these polynomials are closed. In addition, it is known from Erdélyi, Magnus, Oberhettinger, and Tricomi (1954a); Lebedev and Silverman (1965); Szegö (1939) that the series also converges pointwise to f at every continuity point of f if f is piecewise smooth in $\mathcal{I}_w$, i.e.,

$$\lim_{N\to\infty} \sum_{n=0}^{N} \frac{\langle p_n, f\rangle_w}{\|p_n\|_w} p_n(x) = f(x),$$

for every point of continuity x. At a discontinuity point x, it holds that

$$\lim_{N\to\infty} \sum_{n=0}^{N} \frac{\langle p_n, f\rangle_w}{\|p_n\|_w} p_n(x) = \frac{1}{2}\left(\lim_{z\to x^+} f(z) + \lim_{z\to x^-} f(z)\right).$$

Approximations based on OP

Observe that an OP set Π_w associated with a weight function w defined on an interval $\mathcal{I}_w = (a, b)$ can be transformed into an orthonormal set by normalising each element $p_n \in \Pi_w$, i.e., replacing it by $p_n/\|p_n\|_w$. Thus, only orthonormal sets of polynomials are considered in the sequel for the approximation of functions. Let Π_w be one of these, assume it is closed, and denote its elements by p_n.

[5] Appendix A.1 summarised useful properties of classical OPs. The notation #A.1 indicates the item in the appendix A where further details on a particular mathematical notion are given.

Table 2.1: Properties of classical OPs (see, e.g., Abramowitz and Stegun (1965)): Orthogonality intervals, weight functions, and norms. The parameters α, β, and λ satisfy the constraints $\alpha, \beta > -1$, $\lambda > -1/2$, and $\lambda \neq 0$.

Name	p_n	$\mathcal{I}_w$	$w : \tau \mapsto w(\tau)$	$\|p_n\|_w^2$
Jacobi	$\mathrm{P}_n^{(\alpha,\beta)}$	$(-1,1)$	$(1-\tau)^\alpha(1+\tau)^\beta$	$\frac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)}$
Gegenbauer	$\mathrm{C}_n^{(\lambda)}$	$(-1,1)$	$(1-\tau^2)^{\lambda-\frac{1}{2}}$	$\frac{2^{1-2\lambda}\pi\Gamma(n+2\lambda)}{(n+\lambda)(\Gamma(\lambda))^2 n!}$
Chebychev of the first kind	T_n	$(-1,1)$	$(1-\tau^2)^{-\frac{1}{2}}$	$\begin{cases}\frac{1}{2}\pi, & n>0\\ \pi, & n=0\end{cases}$
Chebychev of the second kind	U_n	$(-1,1)$	$(1-\tau^2)^{\frac{1}{2}}$	$\frac{1}{2}\pi$
Chebychev of the third kind	V_n	$(-1,1)$	$(1-\tau)^{-\frac{1}{2}}(1+\tau)^{\frac{1}{2}}$	π
Chebychev of the fourth kind	W_n	$(-1,1)$	$(1-\tau)^{\frac{1}{2}}(1+\tau)^{-\frac{1}{2}}$	π
Legendre	P_n	$(-1,1)$	1	$\frac{2}{2n+1}$
Laguerre	$\mathrm{L}_n^{(\alpha)}$	$(0,\infty)$	$\mathrm{e}^{-\tau}\tau^\alpha$	$\frac{\Gamma(n+\alpha+1)}{n!}$
Hermite	H_n	$(-\infty,\infty)$	$\mathrm{e}^{-\tau^2}$	$\sqrt{\pi}2^n n!$

The best approximation of a function $f \in \mathcal{L}_w^2(\mathcal{I}_w)$ is now studied in detail. Let

$$\mathcal{S}\{f\} = \sum_{n=0}^{\infty} f_n p_n, \quad f_n = \langle f, p_n \rangle_w, \tag{2.8}$$

and

$$\hat{f}_N = \sum_{n=0}^{N} f_n p_n, \tag{2.9}$$

be the Fourier expansion of f and its N-th order truncation, respectively. Recall from Section 2.1.2 that $\hat{f}_N$ is the best approximation of f. The function $\hat{f}_N$ can also be written as

$$\hat{f}_N = \langle f, \mathrm{R}_{N,\tau} \rangle_w, \quad \text{with} \quad \mathrm{R}_{N,\tau} : \sigma \mapsto \mathrm{R}_{N,\tau}(\sigma) = \sum_{n=0}^{N} p_n(\tau) p_n(\sigma). \tag{2.10}$$

Since $p = \langle p, \mathrm{R}_{N,\tau} \rangle_w$, for all $p \in \pi_N$, $\mathrm{R}_{N,\tau}$ is called the reproducing kernel of π_N with respect to Π_w (see, e.g., Luenberger (1997); Meinardus (1967)).

The error $e = \mathcal{S}\{f\} - \hat{f}_N$ stemming from the N-th order truncation of $\mathcal{S}\{f\}$ is analysed in the following theorem[6]. It shows that the zeros of OPs play a crucial role for the approximation error. This will be discussed deeper when causal estimators based on classical OPs are introduced in Section 2.2.

[6]The proof of Theorem 2.3 follows the lines of the proof developed in Kiltz (2017) for the numerical differentiation using Jacobi OPs.

Theorem 2.3: Error from truncation

Let Π_w be a closed orthonormal set of polynomials associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$, $\mathcal{I}_w = (a, b) \subseteq \mathbb{R}$, and $f \in \mathcal{L}^2_w(\mathcal{I}_w)$ be an $m+1$ times differentiable function. Denote by $\mathcal{S}\{f\}$ and $\hat{f}_N$ the Fourier expansion and the best approximation of f given in (2.8) and (2.9), respectively.

The N-th order truncation error $e = \mathcal{S}\{f\} - \hat{f}_N$ evaluated at $\sigma \in \mathcal{I}_w$ satisfies

$$|e(\sigma)| \leqslant \frac{\sup_{\theta \in \mathcal{I}_w} \left| f^{(\rho+1)}(\theta) \right|}{\rho!} M, \tag{2.11}$$

with $\rho = \min\{\bar{N}, m\}$,

$$\begin{aligned} \bar{N} &= \begin{cases} N+1, & \text{for} \quad \sigma \in \{z_n\}_{n=0}^{N+1}, \\ N, & \text{otherwise}, \end{cases} \\ M &= \int_a^b \left| \sum_{n=\bar{N}+1}^{\infty} \langle h_\sigma^\rho, p_n \rangle p_n(\sigma) \right| \mathrm{d}\sigma, \\ h_\sigma^\rho : \theta \mapsto h_\sigma^\rho(\theta) &= \begin{cases} (\sigma - \theta)^\rho, & \text{for} \quad \sigma > \theta, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.12}$$

and z_n the zeros of the polynomial p_{N+1}.

Proof. The truncation error is defined as

$$e(\sigma) = \sum_{n=0}^{\infty} f_n p_n(\sigma) - \sum_{n=0}^{N} f_n p_n(\sigma)$$

for $\sigma \in \mathcal{I}_w$. It satisfies

$$e(\sigma) = \sum_{n=\bar{N}+1}^{\infty} f_n p_n(\sigma),$$

for

$$\bar{N} = \begin{cases} N+1, & \text{for} \quad \sigma \in \{z_n\}_{n=0}^{N+1}, \\ N, & \text{otherwise}, \end{cases}$$

with z_n the zeros of the polynomial p_{N+1}. From Taylor's theorem (see, e.g., Apostol (1967)) it follows that the expansion of f in $\mathcal{I}_w$ around $t_0 \in \mathcal{I}_w$ reads

$$f(\sigma) = \sum_{i=0}^{\rho} \frac{f^{(i)}(t_0)}{i!} (\sigma - t_0)^i + R_\rho(\sigma), \quad \rho \in \{0, \ldots, m\},$$

with the remainder

$$R_\rho(\sigma) = \int_a^b \frac{f^{(\rho+1)}(\theta)}{\rho!} h_\sigma^\rho(\theta) \mathrm{d}\theta,$$

with the function h_σ^ρ defined in (2.12). Since p_n is orthogonal to any polynomial of lower degree, the projection of f onto p_n is

$$f_n = \langle f, p_n \rangle_w = \sum_{i=n}^{\rho} \frac{f^{(i)}(t_0) \langle \pi_i, p_n \rangle_w}{i!} + \int_a^b \frac{f^{(\rho+1)}(\theta) \langle h_\sigma^\rho, p_n \rangle}{\rho!} \mathrm{d}\theta,$$

with $\pi_i : \sigma \mapsto \pi_i(\sigma) = (\sigma - t_0)^i$, for $\rho \geqslant n$. Choosing $\rho = \min\{\bar{N}, m\}$ yields

$$e(\sigma) = \sum_{n=\bar{N}+1}^{\infty} \int_a^b \frac{f^{(\rho+1)}(\theta)\langle h_\sigma^\rho, p_n\rangle}{\rho!} \mathrm{d}\theta p_n(\sigma).$$

Recall that this series converges due to the closure of the orthonormal set considered here. Thus, the summation and the integration in the latter equation can be exchanged such that

$$e(\sigma) = \int_a^b \frac{f^{(\rho+1)}(\theta)}{\rho!} \sum_{n=\bar{N}+1}^{\infty} \langle h_\sigma^\rho, p_n\rangle \mathrm{d}\theta p_n(\sigma).$$

By the mean value theorem it finally follows that

$$|e(\sigma)| \leqslant \frac{\sup_{\theta\in\mathcal{I}_w} \left|f^{(\rho+1)}(\theta)\right|}{\rho!} \int_a^b \left| \sum_{n=\bar{N}+1}^{\infty} \langle h_\sigma^\rho, p_n\rangle p_n(\sigma) \right| \mathrm{d}\theta.$$

□

The parameter M from (2.11) depends only on the properties of the orthonormal set used for the approximation and does not depend on the function to be approximated. From the proof of the theorem it is clear that the roots of the OP p_{N+1} play an import role: Evaluating the approximation at a zero of a specific polynomial decreases the order of the approximation error. This will be even more clear in the next section.

2.2 Causal real-time signal approximation

In numerous applications the approximation of a function

$$\begin{aligned} f : \mathcal{I}(t) &\to \mathbb{R}, \\ \tau &\mapsto f(\tau) \end{aligned} \tag{2.13}$$

in $\mathcal{I}(t) = [0, t]$, for an arbitrary $t > 0$, is sought. Thus, an approximation problem as in Problem 2.1 has to be solved. It shall be assumed in the following that $f \in \mathcal{L}^2(\mathcal{I}(t))$. The approximation has to be evaluated at an arbitrary $\sigma \in [t - t_0, t] \subseteq \mathcal{I}(t)$, $0 \leqslant t_0 \leqslant t$.

The interval $\mathcal{I}(t)$ does not necessarily correspond to the orthogonality interval $\mathcal{I}_w$ of an orthonormal polynomial set Φ_w associated with a weight function w. Thus, the approximation schemes based on truncated Fourier expansions developed earlier cannot immediately be applied to this problem. However, this can be overcome by suitable time transformations and function extensions.

2.2.1 Causal approximation

In the following a bijective function

$$\begin{aligned} \varphi_{t,T} : \mathcal{I}_w &\to \underline{\mathcal{I}(t)}, \\ \tau &\mapsto \varphi_{t,T}(\tau), \end{aligned} \tag{2.14a}$$

is used to map the orthogonality interval $\mathcal{I}_w$ of the orthogonal functions of interest to an arbitrary interval $\underline{\mathcal{I}(t)} \subseteq \mathbb{R}$. It shall be assumed that $\mathcal{I}(t) \subseteq \underline{\mathcal{I}(t)}$. The function $\varphi_{t,T}$ might depend on some parameter T and its inverse is denoted by

$$\begin{aligned} \varphi_{t,T}^{-1} : \underline{\mathcal{I}(t)} &\to \mathcal{I}_w, \\ \tau &\mapsto \varphi_{t,T}^{-1}(\tau). \end{aligned} \tag{2.14b}$$

The function $\varphi_{t,T}$, called time transformation in the sequel, will be defined such that the composition of f, or a suitable extension of it, and $\varphi_{t,T}$ is an element of $\mathcal{L}^2_w(\mathcal{I}_w)$. The properties of the resulting approximation are then studied.

Time transformations

The time transformations are now discussed for closed, half-open, and open orthogonality intervals separately.

Closed intervals: For a closed interval $\mathcal{I}_w = [a, b]$, it is sufficient to find a function mapping the orthogonality interval $\mathcal{I}_w$ to an interval $[t - T, t] = \underline{\mathcal{I}(t)}$, $T \geqslant t_0$. The parameter T might not necessarily be equal to t_0 and is a degree of freedom in the design. The explicit definition of an affine function $\varphi_{t,T}$ from (2.14) performing this transformation is given in Table 2.2. Its effects on the approximation will be discussed in detail in the next section.

Half-open intervals: For a half-open interval[7] $\mathcal{I}_w = [a, \infty)$ the function f can be extended as

$$\begin{aligned} &f_1 : (-\infty, t] \to \mathbb{R}, \\ &\tau \mapsto f_1(\tau) = \begin{cases} f(\tau), & \text{for} \quad \tau \geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.15}$$

Then, f_1 is approximated in the interval $\underline{\mathcal{I}(t)} = (-\infty, t]$ and the approximation is evaluated at $\tau = \sigma$. The explicit definition of an affine function $\varphi_{t,T}$ of the form (2.14) mapping $\mathcal{I}_w$ to $\underline{\mathcal{I}(t)}$ is given in Table 2.2. This corresponds to flipping and scaling the interval $\mathcal{I}_w$. The scaling, here denoted by the parameter T, is a degree of freedom in the design. Its effects on the approximation are discussed in detail in the next section.

Open intervals: For the open interval $\mathcal{I}_w = (-\infty, \infty)$ the function f can be extended as

$$\begin{aligned} &f_1 : (-\infty, \infty) \to \mathbb{R}, \\ &\tau \mapsto f_1(\tau) = \begin{cases} f(\tau), & \text{for} \quad \tau \in \mathcal{I}(t), \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.16}$$

The function f_1 is then approximated on the new interval $\underline{\mathcal{I}(t)} = (-\infty, \infty)$. The explicit definition of an affine function $\varphi_{t,T}$ of the form (2.14) mapping $\mathcal{I}_w$ to $\underline{\mathcal{I}(t)}$ is given in Table 2.2. The scaling and translation parameters T and μ, respectively, are again degrees of freedom. Their effects on the approximation are discussed in the next section.

Summary: All previously discussed time transformations $\varphi_{t,T}$ summarised in Table 2.2 are of the form

$$\varphi_{t,T} : \tau \mapsto \varphi_{t,T}(\tau) = \kappa_0 T \tau + t + \kappa_1 \tag{2.17}$$

with κ_0 and κ_1 depending on the parameters of the orthogonality interval. The function

$$\theta_T : \tau \mapsto \theta_T(\tau) = \varphi^{-1}_{t,T}(t - \tau) \tag{2.18}$$

mapping the point $t - \tau$ to the corresponding point in the orthogonality interval will be useful in later discussions.

[7] A similar procedure can be used for the interval $\mathcal{I}_w = (-\infty, a]$.

Table 2.2: Time transformation $\varphi_{t,T}$ defined in (2.14) mapping the orthogonality interval $\mathcal{I}_w$ to $\underline{\mathcal{I}(t)}$ and its inverse $\varphi_{t,T}^{-1}$. It is assumed that $T, \mu > 0$.

	$\underline{\mathcal{I}(t)}$	$\varphi_{t,T} : \mathcal{I}_w \to \underline{\mathcal{I}(t)}$ $\tau \mapsto \varphi_{t,T}(\tau)$	$\varphi_{t,T}^{-1} : \underline{\mathcal{I}(t)} \to \mathcal{I}_w$ $\tau \mapsto \varphi_{t,T}^{-1}(\tau)$
$\mathcal{I}_w = [a, b]$	$[t-T, t]$	$\varphi_{t,T}(\tau) = t + \frac{T}{b-a}(\tau - b)$	$\varphi_{t,T}^{-1}(\tau) = \frac{b-a}{T}(\tau - t) + b$
$\mathcal{I}_w = [a, \infty)$	$(-\infty, t]$	$\varphi_{t,T}(\tau) = -T(\tau - a) + t$	$\varphi_{t,T}^{-1}(\tau) = \frac{t-\tau}{T} + a$
$\mathcal{I}_w = (-\infty, \infty)$	$(-\infty, \infty)$	$\varphi_{t,T}(\tau) = T\tau - \mu + t$	$\varphi_{t,T}^{-1}(\tau) = \frac{\tau - t + \mu}{T}$

Function approximation using time transformations

Using the former time transformation it follows that $f \circ \varphi_{t,T}^{-1} \in \mathcal{L}_w^2(\mathcal{I}_w)$ when $\mathcal{I}_w$ is closed. Similarly $f_1 \circ \varphi_{t,T}^{-1} \in \mathcal{L}_w^2(\mathcal{I}_w)$ with f_1 given in (2.15) and (2.16) for half-open and open interval $\mathcal{I}_w$, respectively. Then, the approximation schemes developed earlier can be applied. To simplify the notation, f is considered in the sequel and it should be replaced by f_1 whenever half-open or open intervals are considered.

The generalised Fourier transform of f with respect to an orthonormal polynomial set Π_w is

$$\mathcal{S}\{f\} = \sum_{n=0}^{\infty} f_n p_n \circ \varphi_{t,T}^{-1}, \quad f_n = \langle f \circ \varphi_{t,T}, p_n \rangle_w . \tag{2.19}$$

The best approximation $\hat{f}_N$ of f evaluated at the sought point σ is

$$\hat{f}_N(\sigma) = \left\langle f \circ \varphi_{t,T}, \mathrm{R}_{N,\varphi_{t,T}^{-1}(\sigma)} \right\rangle_w \tag{2.20}$$

with the reproducing kernel $\mathrm{R}_{N,\varphi_{t,T}^{-1}(\sigma)}$ defined in (2.10). The following result bounding the error $e = \mathcal{S}\{f\} - \hat{f}_N$ can immediately be derived from the general Theorem 2.3.

Corollary 2.1

Let Π_w be a closed orthonormal set of polynomials associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$, $\mathcal{I}_w \subseteq \mathbb{R}$. Consider an $m+1$ times differentiable function $f : \underline{\mathcal{I}(t)} \to \mathbb{R}$ satisfying $f \circ \varphi_{t,T}^{-1} \in \mathcal{L}_w^2(\mathcal{I}_w)$, with $\varphi_{t,T}$ the time transformation from (2.14) and (2.17) with explicit form given in Table 2.2. Denote by $\mathcal{S}\{f\}$ and $\hat{f}_N$ its Fourier expansion and its best approximation defined in (2.19) and (2.20), respectively.

Then, the N-th order truncation error $e = \mathcal{S}\{f\} - \hat{f}_N$ evaluated at $\sigma \in \underline{\mathcal{I}(t)}$ satisfies

$$|e(\sigma)| \leqslant |\kappa_0 T|^{\rho+1} \frac{\sup_{\theta \in \underline{\mathcal{I}(t)}} \left| f^{(\rho+1)}(\theta) \right|}{\rho!} M = \mathcal{O}\left(T^{\rho+1}\right), \quad \text{for} \quad T \to 0,$$

with $\rho = \min\{\bar{N}, m\}$,

$$\bar{N} = \begin{cases} N+1, & \text{for} \quad \varphi_{t,T}^{-1}(\sigma) \in \{z_n\}_{n=0}^{N+1}, \\ N, & \text{otherwise}, \end{cases}$$

and M defined in (2.11).

From Corollary 2.1, it is clear that the approximation error is reduced when the parameter T of the time transformations is decreased. It is also clear that when the approximation is evaluated at σ with $\varphi_{t,T}^{-1}(\sigma)$ a zero of the orthonormal polynomial of degree $N+1$ the approximation error is reduced by one order.

Observe that any $\sigma \in \underline{\mathcal{I}(t)}$ can be written as $\sigma = t - \delta$, for some δ. The definitions from Table 2.3 of the time transformation yield that $\varphi_{t,T}^{-1}(\sigma) =: \vartheta$ is independent of t. It depends only on δ and the parameters of the time transformation. Thus, if ϑ is a zero of the orthonormal polynomial of degree $N+1$ the order of the approximation error is reduced by one.

2.2.2 Degree of exactness and delay

Evaluating the best approximation $\hat{f}_N$ of a function f at $\sigma = t - \delta \in \underline{\mathcal{I}(t)}$, for some δ, yields the reproducing kernel $\mathrm{R}_{N,\vartheta}$, with $\vartheta := \varphi_{t,T}^{-1}(\sigma)$, defined in (2.10) and given as

$$\mathrm{R}_{N,\vartheta} : \sigma \mapsto \mathrm{R}_{N,\vartheta}(\sigma) = \sum_{n=0}^{N} p_n(\vartheta) p_n(\sigma). \tag{2.21}$$

From the definition of the time transformation $\varphi_{t,T}$ given in Table 2.2 it follows that ϑ depends only on T, δ, and the parameters of the OPs but not on t.

Let $\hat{f}(t) = \hat{f}_N(t-\delta)$, i.e.,

$$\hat{f}(t) = \langle f \circ \varphi_{t,T}, \mathrm{R}_{N,\vartheta} \rangle_w ,$$

and recall that $\mathrm{R}_{N,\vartheta}$ is a reproducing kernel of π_N. The definitions of the time transformation given in Table 2.2 yield

$$\hat{f}(t) = (f \circ \varphi_{t,T})(\vartheta) = f(t - \delta_t), \tag{2.22}$$

for $f \in \pi_N$, where δ_t depends on the choice of the OP set. Thus, the approximation is exact up to a delay δ_t. This observation motivates the definition of the degree of exactness (DoE) initially introduced in Kiltz (2017) and recalled in Othmane et al. (2022) for diffs. based on Jacobi OPs.

Definition 2.4: Degree of Exactness

Let $\mathrm{R}_{N,\vartheta}$, $N \in \mathbb{N}$ and $\vartheta \in \mathbb{R}$, defined in (2.21) be a reproducing kernel of π_N with respect to a closed orthonormal set Π_w of polynomials associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$, with $\mathcal{I}_w \subseteq \mathbb{R}$. The degree of exactness γ of $\mathrm{R}_{N,\vartheta}$ is the polynomial degree up to which the approximation is exact but delayed with a known delay δ_t.

From (2.21) it is clear that if ϑ is a zero of the polynomial p_{N+1} it holds that $\mathrm{R}_{N,\vartheta} = \mathrm{R}_{N+1,\vartheta}$. Then, even for $f \in \pi_{N+1}$ the relation (2.22) is valid. This corresponds to the result from Corollary 2.1 stating that the estimation error is reduced by one order for special choices of the evaluation time instant. Since $p_0 \equiv 1$, the reproducing kernel $\mathrm{R}_{0,\vartheta}$ is independent of ϑ and it holds that

$$\mathrm{R}_{0,\vartheta} = \mathrm{R}_{1,z_1}, \tag{2.23}$$

with z_1 the zero of the polynomial p_1. These observation can be summarised in the following result for the DoE.

Corollary 2.2: Degree of exactness

Let $\mathrm{R}_{N,\vartheta}$, $\vartheta \in \mathbb{R}$ and $N \in \mathbb{N}$, defined in (2.21) be a reproducing kernel of π_N with respect to a closed orthonormal set Π_w of polynomials associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$, with $\mathcal{I}_w \subseteq \mathbb{R}$. The degree of exactness $\gamma_{\mathrm{R}_{N,\vartheta}}$ of $\mathrm{R}_{N,\vartheta}$ is

$$\gamma_{\mathrm{R}_{N,\vartheta}} = \begin{cases} N+k, & \text{for} \quad N = 0 \vee \vartheta \in \{z_n\}_{n=0}^{N+1}, \\ N, & \text{otherwise}, \end{cases}$$

with $\{z_n\}_{n=0}^{N+1}$ the zeros of the polynomial p_{N+k}.

In numerous applications the function f needs to be approximated at $\sigma = t$, i.e., at the upper boundary of its definition interval. It follows from (2.23) that

$$\hat{f}(t) = f(t - \delta_t), \quad \delta_t = z_1,$$

for $N = 0$ and $f \in \pi_1$. Thus, for $N = 0$ the approximation of f at $\sigma = t$ is independent of ϑ and necessarily delayed. For $N > 0$, it follows from Table 2.2 that

$$\delta_t = \bar{\theta}_T(\vartheta) = t - \varphi_{t,T}(\vartheta) \tag{2.24}$$

and δ_t depends only on the parameters ϑ, T, and the orthonormal set chosen for the approximation. By adapting ϑ the parametrisation $\delta_t = 0$ is possible, i.e., a delay-free approximation is achieved. For this reason, in the remaining parts of this work the quantity δ_t is called the estimation delay of the reproducing kernel $\mathrm{R}_{N,\vartheta}$. Its analytic representation for classical OPs is recalled in Table 2.3. These observations are now summarised in the following corollary.

Corollary 2.3: Estimation delay

Let $\mathrm{R}_{N,\vartheta}$, $\vartheta \in \mathbb{R}$, defined in (2.21) be a reproducing kernel of π_N with respect to a closed orthonormal set Π_w of polynomials associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$, with $\mathcal{I}_w \subseteq \mathbb{R}$. The estimation delay δ_t of $\mathrm{R}_{N,\vartheta}$ is

$$\delta_t = \begin{cases} \bar{\theta}(z_1), & \text{for} \quad N = 0, \\ \bar{\theta}_T(\vartheta), & \text{otherwise}, \end{cases}$$

where z_1 is the zero of the polynomial $p_1 \in \Pi_w$ and $\bar{\theta}_T$ is given in (2.24).

In summary, for $N = 0$ the best approximation of f is necessarily delayed. By varying the parameter ϑ of the reproducing kernel, an arbitrary value can be assigned to the estimation delay δ_t. Accepting a small but known delay increases the degree of exactness and decreases the approximation error by one order. The delay is minimised by choosing the appropriate zero of the corresponding OP. Recalling from Theorem 2.1 that the zeros of OPs are distinct and lie within the interior of the orthogonality interval of the polynomials, it follows that for closed intervals the estimation delay also lies in the interval $(t - T, t)$. Using Theorem 2.2 it can also be concluded that the estimation delay can be decreased by increasing N.

Table 2.3: Delay δ_t from Corollary 2.3 for $N > 1$ computed for classical OPs.

	Jacobi	Laguerre	Hermite
δ_t	$\frac{1-\vartheta}{2}T$	$T\vartheta$	$\mu - T\vartheta$

2.3 The approximation problem of derivatives of arbitrary orders

In the preceding section the causal approximation of a function

$$\begin{aligned} f : \mathcal{I}(t) &\to \mathbb{R}, \\ \tau &\mapsto f(\tau) \end{aligned} \tag{2.25}$$

in $\mathcal{I}(t) = [0, t]$, for an arbitrary $t > 0$, has been discussed. Thus, an approximation problem as in Problem 2.1 has to been solved.

The results of the last section can be immediately extended for the estimation of derivatives of arbitrary large order. Thus, assuming that f is sufficiently smooth, the estimation of $f^{(n)}$, the n-th order derivative of f, is considered in this section. In the sequel, it is also assumed that only an additively disturbed measurement $\tilde{f}$ of f is available, i.e.,

$$\tilde{f} : t \mapsto \tilde{f}(t) = f(t) + \eta(t), \tag{2.26}$$

where η is a bounded disturbance.

This section is divided into four parts. Application of the previous approaches for the approximation of the derivatives $f^{(n)}$ using only $\tilde{f}$ is first discussed and the different errors are analysed. Then, the derivation of these methods using differential-algebraic manipulations of Taylor series is analysed. Finally, system-theoretic, wavelet-based and a regularisation-based interpretations are provided.

2.3.1 An approximation-theoretic point of view

Using the methods developed in Section 2.2 the following approximation schemes can be immediately derived[8]. Recall that the best approximation $\hat{f}_N^{(n)}$ evaluated at $\sigma = t - \delta \in \underline{\mathcal{I}(t)}$, for some δ, of $f^{(n)}$ is

$$\hat{f}_N^{(n)}(t-\delta) = \left\langle f^{(n)} \circ \varphi_{t,T}, \mathrm{R}_{N,\vartheta} \right\rangle_w, \quad \vartheta = \theta_T(\delta), \tag{2.27}$$

where the reproducing kernel $\mathrm{R}_{N,\vartheta}$ is defined in (2.21) and θ_T in (2.18). Let

$$g_{N,\vartheta} : \tau \mapsto g_{N,\vartheta}(\tau) = w(\tau)\mathrm{R}_{N,\vartheta}(\tau). \tag{2.28}$$

Using the definition of the time transformation in (2.17), a repeated integration by parts of (2.27) yields[9]

$$\begin{aligned} \hat{f}_N^{(n)}(t-\delta) &= \int_a^b g_{N,\vartheta}(\tau) \left(f^{(n)} \circ \varphi_{t,T} \right)(\tau) \mathrm{d}\tau \\ &= \frac{1}{\kappa_0^n T^n} \int_a^b g_{N,\vartheta}(\tau) \left(f \circ \varphi_{t,T} \right)^{(n)}(\tau) \mathrm{d}\tau \\ &= \frac{1}{\kappa_0^n T^n} \left[\sum_{m=0}^{n-1} (-1)^m g_{N,\vartheta}^{(m)}(\tau) f^{(n-1-m)}(\tau) \right]_a^b + \frac{(-1)^n}{\kappa_0^n T^n} \int_a^b \left(f \circ \varphi_{t,T} \right)(\tau) g_{N,\vartheta}^{(n)}(\tau) \mathrm{d}\tau. \end{aligned}$$

[8] When an open or half-open orthogonality interval $\mathcal{I}_w$ is used the function $f^{(n)}$ needs to be replaced in the following by the extended function as discussed in detail in Section 2.2

[9] For a function f and $a = \infty$ the notation $f(a)$ should be read as $f(a) = \lim_{\tau \to \infty} f(\tau)$

Thus, if the first $n-1$ derivatives of $g_{N,\vartheta}$ vanish at the boundaries of the orthogonality interval $\mathcal{I}_w = (a,b)$, the best approximation $\hat{f}_N^{(n)}$ evaluated at $t-\delta$ of $f^{(n)}$ can be computed using solely f as

$$\hat{f}_N^{(n)}(t-\delta) = \frac{(-1)^n}{\kappa_0^n T^n} \int_a^b \left(f \circ \varphi_{t,T}\right)(\tau) g_{N,\vartheta}^{(n)}(\tau) \mathrm{d}\tau. \tag{2.29}$$

This observation yields the definition of an orthogonal differentiator (orth. diff.) used in the remaining parts of this work.

Definition 2.5: Orthogonal differentiator

Let $\mathrm{R}_{N,\vartheta}$, $\vartheta \in \mathbb{R}$ and $N \in \mathbb{N}$ finite, defined in (2.21) be a reproducing kernel of π_N with respect to a closed orthonormal set Π_w of polynomials associated with a weight function $w : \mathcal{I}_w \to \mathbb{R}$, with $\mathcal{I}_w \subseteq \mathbb{R}$. Assume that the first $n-1$ derivatives of $g_{N,\vartheta}$ defined in (2.28) vanish at the boundaries of $\mathcal{I}_w$. The function $g_{N,\vartheta}$ is called an orthogonal diff. of order n associated with Π_w.

Replacing in (2.29) the function f by the measurement $\tilde{f}$ yields the approximation

$$\hat{\tilde{f}}_N^{(n)}(t-\delta) = \frac{(-1)^n}{\kappa_0^n T^n} \int_a^b \left(\tilde{f} \circ \varphi_{t,T}\right)(\tau) g_{N,\vartheta}^{(n)}(\tau) \mathrm{d}\tau$$

of $f^{(n)}$. The latter approximation requires only the measurement $\tilde{f}$. The concept of differentiation by integration can be clearly seen here. Approaches for the approximation of derivatives using integrals have been known for a long time as can be seen in (Lanczos, 1956, P. 324) and Cioranescu (1938), for example.

The approximated derivative $\hat{\tilde{f}}_N^{(n)}(t-\delta)$ is corrupted by two sources of errors[10]:

1. the bias term, denoted by e_N, stemming from the truncation of the Fourier expansion and
2. the variance term, denoted by e_η, stemming from the disturbance η.

The bias term $e_N = \mathcal{S}\{f^{(n)}\} - \hat{f}_N^{(n)}$ can be immediately bounded using Corollary 2.1 to get the following result.

Corollary 2.4

Let $g_{N,\vartheta}$ be an orth. diffs. associated with Π_w a closed orthonormal set of polynomials. Consider an $n+m+1$ times differentiable function $f : \underline{\mathcal{I}(t)} \to \mathbb{R}$ satisfying $f \circ \varphi_{t,T} \in \mathcal{L}_w^2(\mathcal{I}_w)$, with $\varphi_{t,T}$ the time transformation from (2.14) and (2.17) with explicit forms given in Table 2.2. Denote by $\mathcal{S}\{f^{(n)}\}$ and $\hat{f}_N^{(n)}$ the Fourier expansion of $f^{(n)}$ and its best approximation defined in (2.19) and (2.29), respectively.

Then, the N-th order truncation error $e_N = \mathcal{S}\{f^{(n)}\} - \hat{f}_N^{(n)}$ evaluated at $t-\delta \in \underline{\mathcal{I}(t)}$ satisfies

$$|e_N(t-\delta)| \leqslant |\kappa_0 T|^{\rho+1} \frac{\sup_{\theta \in \underline{\mathcal{I}(t)}} \left|f^{(n+\rho+1)}(\theta)\right|}{\rho!} M = \mathcal{O}\left(T^{\rho+1}\right), \quad \text{for} \quad T \to 0,$$

with ρ and M defined in Corollary 2.1.

[10] The nomenclature of the errors follows the lines of Mboup et al. (2009).

The variance term $e_\eta = \hat{\tilde{f}}^{(n)} - \hat{f}^{(n)}$ reads

$$\begin{aligned} e_\eta(t-\delta) &= \frac{(-1)^n}{\kappa_0^n T^n}\left(\int_a^b \left(\left(\tilde{f}-f\right)\circ\varphi_{t,T}\right)(\tau) g_{N,\vartheta}^{(n)}(\tau)\mathrm{d}\tau\right) \\ &= \frac{(-1)^n}{\kappa_0^n T^n}\left(\int_a^b (\eta\circ\varphi_{t,T})(\tau) g_{N,\vartheta}^{(n)}(\tau)\mathrm{d}\tau\right) \end{aligned} \tag{2.30}$$

and it is straightforward to see that it satisfies the bound in the following corollary.

Corollary 2.5: Variance term

Let f and η be two functions satisfying $f\circ\varphi_{t,T}, \eta\circ\varphi_{t,T} \in \mathcal{L}_w^2(\mathcal{I}_w)$ and $\tilde{f}$ be the function defined in (2.26). Assume that η is bounded and $f^{(n)}\circ\varphi_{t,T} \in \mathcal{L}_w^2(\mathcal{I}_w)$. The variance term e_η defined in (2.30) satisfies $|e_\eta(t-\delta)| = \mathcal{O}(T^{-n})$ for $T \to 0$.

From Corollaries 2.4 and 2.5 a known dilemma in function approximation can be observed: Reducing the parameter T reduces the bias term but significantly increases the variance term. For example, consider a closed orthogonality interval. The parameter T is then the width of the approximation window. For a fixed N decreasing the window length T will decrease the bias term $|e_N(t-\delta)|$. However, a large T is required to smooth out the effects of the disturbance and thus to reduce the variance term $|e_\eta(t-\delta)|$. Hence, finding a compromise is important.

As for reproducing kernels, an orth. diff. has a DoE $\gamma_{g_{N,\vartheta}}$, i.e., the polynomial degree up to which the derivative approximation is exact but delayed with a known delay. It follows immediately from Corollary 2.2 as $n + \gamma_{\mathrm{R}_{N,\vartheta}}$ with $\gamma_{\mathrm{R}_{N,\vartheta}}$ the DoE of the corresponding reproducing $\mathrm{R}_{N,\vartheta}$. The estimation delay corresponds to that of $\mathrm{R}_{N,\vartheta}$ given in Corollary 2.3.

In addition to this approximation theoretic derivation of orth. diffs., an algebraic interpretation can be given in terms of a parameter estimation problem with annihilation of undesired structured perturbations as discussed in the next section.

2.3.2 An algebraic point of view

Let $f : t \mapsto f(t)$ be analytic in $\mathcal{I} \subseteq \mathbb{R}$, with $0 \in \mathcal{I}$, and consider its convergent Taylor series expansion at $t = 0$ namely

$$f(t) = \sum_{m=0}^{\infty} f^{(m)}(0)\frac{t^m}{m!}.$$

For the estimation of the n-th order derivative of f, the truncated Taylor series expansion

$$f_{n^\star}(t) = \sum_{m=0}^{n^\star} f^{(m)}(0)\frac{t^m}{m!} \tag{2.31}$$

for $n^\star \geqslant n$ is considered. It satisfies the linear differential equation $f_{n^\star}^{(n^\star+1)}(t) = 0$, which using operational calculus (Laplace transform or Mikusinski's operational calculus detailed in Doetsch (1974) and Mikusinski (1983), respectively) reads

$$\mathrm{s}^{n^\star+1}\mathcal{L}\{f_{n^\star}\}(\mathrm{s}) = \sum_{m=0}^{n^\star} \mathrm{s}^{n^\star-m} f^{(m)}(0), \tag{2.32}$$

where $\mathcal{L}\{f_{n^\star}\}$ is the operational counterpart of $f_{n^\star}$. It should be noted here that all derivatives $f^{(m)}(0)$, $m \in \{0,\ldots,n^\star\}$ are linearly identifiable parameters of (2.32) according to the identifiability definition developed in Fliess and Sira-Ramírez (2003a) using module theory.

Annihilators

For the estimation of the n-th order derivative of f all terms $s^{n^\star-m}x^{(m)}(0)$, $m \neq n$, i.e., derivatives that are not of interest, are now viewed as perturbations to be annihilated. This can be achieved using differential operators of the form

$$\mathcal{D} = \sum_{l=0}^{l^*} \mathcal{L}\{d_l\}(\mathrm{s}) \left(\frac{\mathrm{d}}{\mathrm{ds}}\right)^l, \quad \mathcal{L}\{d_l\} : \mathbb{C} \to \mathbb{C}, \quad l^* \text{ finite}, \tag{2.33a}$$

satisfying

$$\mathcal{D}\mathcal{L}\{f_N\}(\mathrm{s}) = \mathcal{L}\{d\}(\mathrm{s}) f^{(n)}(0), \tag{2.33b}$$

for some function $\mathcal{L}\{d\} : \mathbb{C} \to \mathbb{C}$. As in Mboup et al. (2009), a differential operator of the form (2.33a) is called an annihilator for $f^{(n)}(0)$. Obviously, such operators exist and are not unique. However, it is clear that to each annihilator $\mathcal{D}$, there is a unique $\mathcal{L}\{d\} : \mathbb{C} \to \mathbb{C}$ such that (2.33b) is satisfied. The operator $\mathcal{D}$ and the function $\mathcal{L}\{d\}$ are said to be associated.

The operator $\mathcal{D}$ has to be judiciously chosen to ensure that only integral operators are involved when the relations are interpreted in the time domain. In the sequel, an operator satisfying this condition is called a proper differential operator in accordance with the discussions in Mboup et al. (2009). It has been shown in the latter reference that for $\mathcal{L}\{d_l\} \in \mathbb{C}(\mathrm{s})$ the operator $\mathcal{D}$ is proper if, and only if, each $\mathcal{L}\{d_l\}$ is a proper rational function[11].

Time-domain approximation

Replacing in (2.33) the truncated Taylor series expansion $f_{n^\star}$ by the actual known signal $\tilde{f}$ from (2.26) yields

$$\mathcal{L}\{d\}(\mathrm{s}) f_{n^\star}^{(n)}(0) = \mathcal{D}\mathcal{L}\left\{\tilde{f}\right\}(\mathrm{s}).$$

This equation provides an operational estimator for $f^{(n)}(0)$, observed by $\tilde{f}$. The latter expression needs now to be transformed in the time domain. Let d and d_l be the time-domain counterparts of $\mathcal{L}\{d\}$ and $\mathcal{L}\{d_l\}$, respectively. Writing the latter equation in the time domain yields[12]

$$d(t)\tilde{f}_{n^\star}^{(n)}(0) = \sum_{l=0}^{l^*} \int_0^t d_l(t-\tau)(-1)^l \tau^l \tilde{f}(\tau)\mathrm{d}\tau. \tag{2.34}$$

Here, the variable t represents the estimation time and the equation has to be considered for fixed t. A finite and an infinite estimation time are considered in the sequel. The relation to truncated Fourier series is discussed.

Finite estimation time

In Mboup et al. (2009) a finite estimation time $t^* > 0$ has been used in (2.34). This yields the approximated derivative

$$\tilde{f}_{n^\star}^{(n)}(0) = \frac{1}{d(t^*)} \sum_{l=0}^{l^*} \int_0^{t^*} d_l(t^*-\tau)(-1)^l \tau^l \tilde{f}(\tau)\mathrm{d}\tau.$$

[11] A rational function is called proper if, and only if, the degree of the nominator is less or equal to that of the denominator.

[12] Recall from the classical rules of operational calculus that $\frac{\mathrm{d}}{\mathrm{ds}}$ in the operational domain corresponds to a multiplication by $-t$ in the time domain (see, e.g., Doetsch (1974); Mikusinski (1983)).

A causal approximation of the derivative denoted in the following by $\hat{f}^{(n)}$ of the n-th order derivative is obtained by replacing in the last equation $f(\tau)$ by $-\Theta(\tau)f(t-\tau)$, for $t > t^*$. The approximated derivative

$$\hat{f}^{(n)}(t) = \int_0^{t^*} \sum_{l=0}^{l^*} \frac{d_l(t^*-\tau)(-1)^{l+1}\tau^l}{d(t^*)} \tilde{f}(t-\tau)\mathrm{d}\tau$$

is then obtained. Observing that this corresponds to an approximation in the sliding window $[t-t^*, t]$ motivates the definition of annihilators that yield a kernel

$$\mathrm{R}(\tau) = \sum_{l=0}^{l^*} \frac{d_l(t^*-\tau)(-1)^l\tau^l}{d(t^*)},$$

which is a reproducing kernel of $\pi_{n^\star-n}$ derived using polynomials orthogonal on a closed interval that can be mapped to the interval $[0, t^*]$ using the time transformation given in Table 2.2.

In Mboup et al. (2009) annihilators that yield a reproducing kernel based on Jacobi polynomials have been derived. These are recalled in Section 2.4.3, where orth. diff. based on classical OP are discussed.

Infinite estimation time

Assuming that the quantity

$$\tilde{f}_{n^\star}^{(n)}(0) = \lim_{t^*\to\infty} \frac{1}{d(t^*)} \sum_{l=0}^{l^*} \int_0^{t^*} d_l(t^*-\tau)(-1)^l\tau^l\tilde{f}(\tau)\mathrm{d}\tau$$

exists and is finite, an infinite estimation time can be used. Let

$$d_\infty(\tau) = \lim_{t^*\to\infty} \frac{1}{d(t^*)} \sum_{l=0}^{l^*} d_l(t^*-\tau)(-1)^l.$$

A causal approximation of the n-th order derivative is obtained by replacing in the last equation $f(\tau)$ by $-\Theta(\tau)f(t-\tau)$, for $t > 0$. The approximated derivative, denoted in the following by $\hat{f}^{(n)}$, is

$$\hat{f}^{(n)}(t) = -\int_0^{\infty} d_\infty(\tau)\tau^l\tilde{f}(t-\tau)\mathrm{d}\tau$$

is then obtained, which corresponds to an approximation in the growing window $[-\infty, t]$.

In Ushirobira (2018) annihilators that yield reproducing kernels based on Laguerre and Hermite polynomials have been derived. These are recalled and generalised Section 2.4.2, for example, where orth. diff. based on Laguerre OP are discussed.

2.3.3 A system-theoretic point of view

In addition to the approximation-theoretic and algebraic interpretations of the approximation scheme derived above, a system-theoretic one can be provided by reconsidering the approximation from (2.29).

The approximation $\hat{f}_N^{(n)}$ defined in (2.29) can be rewritten as

$$\hat{f}_N^{(n)}(t-\delta) = \frac{1}{\kappa_0 T} \int_{\varphi_{t,T}(a)}^{\varphi_{t,T}(b)} f(\tau)\left(g_{N,\vartheta}\circ\theta_T\right)^{(n)}(t-\tau)\mathrm{d}\tau, \tag{2.35}$$

with θ_T given in (2.18). It is clear that orth. diffs. can be interpreted as linear time-invariant filters acting on the measured signals. It is clear that this numerical differentiation approach falls within the framework of classical approaches for differentiation in the frequency domain, where a filter is followed by an ideal diff.

The result of the integration by parts of the latter equation yields

$$\hat{f}_N^{(n)}(t-\delta) = \frac{(-1)^n}{\kappa_0 T} \int_{\varphi_{t,T}(a)}^{\varphi_{t,T}(b)} f^{(n)}(\tau) \left(g_{N,\vartheta} \circ \theta_T\right)(t-\tau) \mathrm{d}\tau. \tag{2.36}$$

Thus, the approximated derivative is the output of a linear time-invariant filter applied on the sought derivative $f^{(n)}$. This interpretation can be used for the parametrisation of the orth. diffs. to achieve desired filter characteristics.

These two interpretations motivate the analysis of the filter characteristics of orth. diff. to derive systematic tuning guidelines.

2.3.4 An inverse-problem point of view

It is well known[13] that the n-th order derivative of an n-times differentiable signal $x : t \mapsto x(t)$ can be seen as a solution of a Fredholm integral equation (see, e.g., Cullum (1971); Engl, Hanke, and Neubauer (1996)). In fact it is an ill-posed inverse problem and the mollification approach (see, e.g., Hegland and Anderssen (1998); Murio (2011)), which is a regularisation technique, is one of numerous numerical methods that have been developed to obtain stable algorithms robust to additive noise. In the control community, these methods have been discussed in Diop, Grizzle, and Chaplais (2000); Diop and Martínez-Guerra (2001), for example.

From Hegland and Anderssen (1998) it follows that a real function m_γ is called a *mollifier* if there exists a constant integer $\nu \in \mathbb{N}$ such that

$$m_\gamma(t) = \frac{1}{\gamma} m\left(\frac{t}{\gamma}\right), \quad m \in \mathcal{C}^\nu(\mathbb{R}),\ t \in \mathbb{R},$$

with the *mollifier's generator* m satisfying the following conditions

$$\begin{aligned} m(t) &\geqslant 0, \quad t \in \mathbb{R}, \\ m^{(k)}(t) &\in \mathcal{L}_1(\mathbb{R}), \quad k = 0, \ldots, s, \quad s \in \mathbb{N}, \\ \int_{-\infty}^{\infty} m(t) \mathrm{d}t &= 1, \\ \int_{-\infty}^{\infty} t^k m(t) \mathrm{d}t &\in \mathbb{R}, \quad k = 0, 1, 2, \ldots \end{aligned} \tag{2.37}$$

where $\mathcal{L}_p(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R} \,|\, \int_{-\infty}^{\infty} |f(x)|^p \,\mathrm{d}x < \infty \right\}$, for $0 < p < \infty$. This is a generalises the definition of Friedrichs' mollifiers discussed in Adams and Fournier (1975) in the context of function approximation.

It is now straightforward to verify that minimal diffs., i.e., those with $N = 0$, are mollifiers, and from Hegland and Anderssen (1998) it follows that differentiation based on orthogonal polynomials with $N = 0$ is a regularisation strategy. Affine diffs., however, do not satisfy the non-negativity condition (2.37) in general. The interpretation of diffs. based on Jacobi OPs as mollifiers has been used in Riachy et al. (2016) instead of a high-gain observer for the differentiation of the output signal in the context of universal integral control introduced in Khalil (2000).

[13] This subsection stems literally from Othmane et al. (2022).

2.3.5 A wavelet-decomposition point of view

Recall the approximation of the n-th order derivative of a function f defined in (2.36) and define

$$h : \tau \mapsto h(\tau) = \begin{cases} \frac{(-1)^n}{\kappa_0\sqrt{T}} \left(g_{N,\vartheta} \circ \theta_T\right)^{(n)} (-\tau T), & \text{for} \quad \tau \in [\varphi_{t,T}(a), \varphi_{t,T}(b)], \\ 0, & \text{otherwise.} \end{cases} \tag{2.38}$$

Then, (2.35) can be written as

$$\hat{f}_N^{(n)}(t-\delta) = \int_{-\infty}^{\infty} f(\tau) \frac{1}{\sqrt{T}} h\left(\frac{\tau - t}{T}\right) \mathrm{d}\tau =: W_h f(t, T),$$

i.e., the approximated derivative is the wavelet decomposition of f, at time t and scale T, denoted by $W_h f(t, T)$ (see, e.g., Mallat (1999)).

Let $\mathcal{F}\{h\}$ be the Fourier transform of h and assume that[14]

$$C_h = \int_0^{\infty} \frac{\left|\mathcal{F}\{h\}(\omega)\right|^2}{\omega} \mathrm{d}\omega < \infty. \tag{2.39}$$

According to (Mallat, 1999, P. 124), this condition is called the wavelet admissibility condition. Then, from the Calderón-Grossman-Morlet Theorem (see, e.g., (Mallat, 1999, Thm. 4.3)) it follows that any function $f \in \mathcal{I}_2$ can be reconstructed from its n-th order derivative approximation $W_h f(\cdot, T)$ with continuously varying parameter $T = \sigma$ by

$$f(t) = \frac{1}{C_h} \int_0^{\infty} \int_{-\infty}^{\infty} W_h f(\tau, \sigma) \frac{1}{\sqrt{\sigma}} h\left(\frac{\tau - t}{\sigma}\right) \mathrm{d}\tau \frac{\mathrm{d}\sigma}{\sigma^2}.$$

This wavelet-based interpretation of numerical differentiation using Jacobi polynomials has first been proposed in Mboup and Riachy (2018).

2.4 Numerical differentiation using classical orthogonal polynomials

The classical OPs have already been introduced in Section 2.1.3 as a particular subset of OPs. The orthonormal sets formed by these polynomials can now be used to approximate functions and their derivatives. The estimation of arbitrary derivatives of a function f defined in (2.25) is considered using these polynomials in the sequel. The disturbance-free case is considered for notational simplicity only since the effects of the additive disturbances in (2.26) have been analysed in detail for the general case, and the results are immediately transferable to the diffs. based on classical OPs.

The time transformations from (2.14) required for adaptation of the orthogonality intervals, the corresponding reproducing kernels, and orth. diffs. defined in (2.9) and Definition 2.5, respectively, can immediately be derived from the previous considerations in Tables 2.1 and 2.2 and (2.9). Thus, it is assumed in the following that the considered functions always satisfy the required integrability conditions used in Section 2.3.1.

The properties of orth. diffs. based on the Hermite, Laguerre, and Jacobi OP are investigated. The time-domain, frequency-domain properties, and the variation of the estimation delays as functions of the involved parameters are discussed. For the diffs. based on Jacobi polynomials, a brief overview describing their different historical derivations summarising the

[14] It is straightforward to show that this condition is satisfied for the diffs. based on classical OPs discussed in Section 2.4. Indeed, as discussed in (Mallat, 1999, P. 124), it is sufficient to see that in this special case $\mathcal{F}\{h\}$ is continuously differentiable and satisfies $\mathcal{F}\{h\}(0) = 0$.

survey in Othmane et al. (2022) is provided. A MATLAB and Python toolbox implementing all necessary functions for the design, analysis, and discretization of Jacobi differentiators (Jac. diffs.) has been made available in Othmane (2021). Its use is demonstrated in the appendix of Othmane et al. (2022) and in the examples provided in Othmane (2021). This publicly available code has been used for the generation of all results relying on Jac. diff. in the current work.

2.4.1 Hermite differentiators

Recall from Table 2.1 that Hermite polynomials are orthogonal on the interval $\mathcal{I}_w = (-\infty, \infty)$ and their weight function w satisfies $\lim_{\tau\to-\infty} w(\tau) = \lim_{\tau\to\infty} w(\tau) = 0$. Thus, the effects of the weight function vanish for large absolute values of the argument and the interval has to be shifted using a parameter μ to ensure that the maximum of w is in an interval of interest. A second degree of freedom can be incorporated by adding a scaling parameter T to compress or dilate the orthogonality interval. The transformation follows from Table 2.2 as

$$\begin{aligned} \varphi_{t,T} :\ & \mathcal{I}_w \to \underline{\mathcal{I}(t)}, \\ & \tau \mapsto \varphi_{t,T}(\tau) = T(\tau - \mu) + t, \end{aligned} \tag{2.40a}$$

and its inverse is

$$\begin{aligned} \varphi_{t,T}^{-1} :\ & \underline{\mathcal{I}(t)} \to \mathcal{I}_w, \\ & \tau \mapsto \varphi_{t,T}^{-1}(\tau) = \frac{\tau - t}{T} + \mu, \end{aligned} \tag{2.40b}$$

where $\mathcal{I}_w = (-\infty, \infty)$ and $\underline{\mathcal{I}(t)} = (-\infty, \infty)$.

The approximation of a function $f : (-\infty, t] \to \mathbb{R}$ is now considered using Hermite polynomials (see Table 2.1 for a summary of their properties). This can be performed using the filter interpretation of the approximation scheme given in (2.35), the definitions (2.18), (2.21), and (2.28). Then, the delayed approximation is

$$\hat{f}(t) = \int_{-\infty}^{t} f(\tau) g_{N,T,\vartheta,\mu}(t - \tau) \mathrm{d}\tau$$

with[15]

$$g_{N,T,\vartheta,\mu} = \frac{1}{T} \sum_{n=0}^{N} \frac{\mathrm{H}_n(\vartheta)}{\|\mathrm{H}_n\|_w} (w \cdot \mathrm{H}_n) \circ \theta_T$$

where

$$\|\mathrm{H}_n\|_w = \sqrt{\pi} 2^n n!, \quad w(\tau) = \mathrm{e}^{-\tau^2}, \quad \theta_T(\tau) = \frac{\mu - \tau}{T} \quad \text{and} \quad \vartheta \in \mathbb{R}.$$

The parameter ϑ influences the estimation delay and the degree of exactness as discussed in Corollary 2.2 and Corollary 2.3. Choosing ϑ as a zero of the Hermite OP of degree $N + 1$ decreases the approximation error by one order as discussed in the Corollaries 2.1 and 2.4 and yields a delayed approximation. The estimation delay is given in Table 2.3 and its variation with respect to the parameter N is shown in Fig. 2.1. As given in Theorem 2.1 increasing N decreases the delay. A delay-free approximation is achieved by choosing $\vartheta = \mu/T$.

Since the orthogonality interval of the Hermite polynomials is $\mathcal{I}_\mathrm{H} = (-\infty, \infty)$, the latter approximation corresponds to that of the extended function

$$\bar{f} : \tau \mapsto \bar{f}(\tau) = \begin{cases} \bar{f}(\tau), & \text{for} \quad \tau \leqslant t \\ 0, & \text{otherwise.} \end{cases}$$

[15] The pointwise product of two functions f and g is denoted by $f \cdot g$.

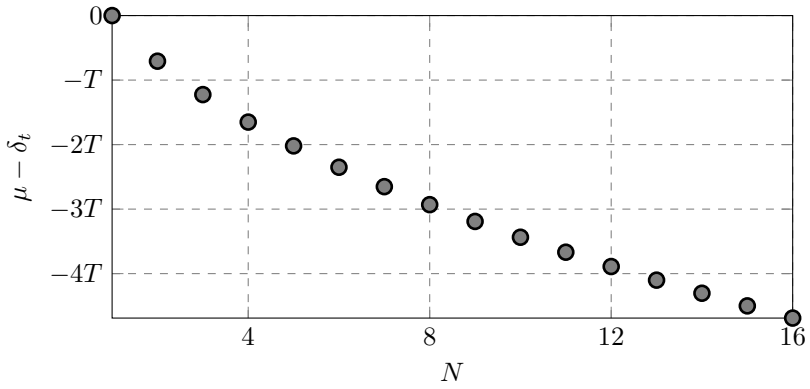


Figure 2.1: Variation of the estimation delay of the Hermite diff. as a function of the parameter N.

Since $\bar{f}$ is discontinuous at time $\tau = t$ the Fourier expansion of f at $\tau = t$ converges to $\bar{f}(t)/2$. Thus, if an approximation at $\tau = t$ is sought the extension should be adapted such that $\bar{f}$ is continuous at this point.

For the approximation of the n-th order derivative of a sufficiently differentiable functions f it has been shown in Section 2.3.1 that the $n-1$ first derivatives of the kernel $g_{N,T,\vartheta,\mu}$ must vanish at the boundaries of $\mathcal{I}_w$. The analytical form of the derivatives of the Hermite polynomials from #A.3 yields

$$\left(g_{N,T,\vartheta,\mu}\right)^{(n)} = \frac{1}{T^{n+1}} \sum_{m=0}^{N} \frac{\mathrm{H}_m\left(\vartheta\right)}{\left\|\mathrm{H}_m\right\|_w} \left(w \cdot \mathrm{H}_{m+n}\right) \circ \theta_T, \tag{2.41}$$

and it is clear that the derivatives of the kernel always vanish as required for a diff.

Definition 2.6: Hermite diff.

Let $N \in \mathbb{N}$ and $T, \mu, \vartheta \in \mathbb{R}$, with $T, \mu > 0$. The function

$$g_{N,T,\vartheta,\mu} : \tau \mapsto g_{N,T,\vartheta,\mu}(\tau) = \frac{1}{T} \sum_{m=0}^{N} \frac{\mathrm{H}_m\left(\vartheta\right)}{\left\|\mathrm{H}_m\right\|_w} \left(\left(w \cdot \mathrm{H}_m\right) \circ \theta_T\right)(\tau), \tag{2.42a}$$

with

$$\left\|\mathrm{H}_m\right\|_w = \sqrt{\pi} 2^m m!, \quad w(\tau) = \mathrm{e}^{-\tau^2} \quad \text{and} \quad \theta_T(\tau) = \frac{\mu - \tau}{T}, \tag{2.42b}$$

is called a Hermite diff. Its n-th order derivative is given in (2.41).

Filter characteristics

Since the approximation of functions and their derivatives using Hermite OPs can be interpreted as filtering the measurement, the properties of these filters are now discussed both in the time domain and the frequency domain.

Time-domain properties: The explicit representation of the impulse response of a Hermite differentiator (Hermite diff.) $g_{N,T,\vartheta,\mu}$ defined in Definition 2.6 is given in (2.42) and can be

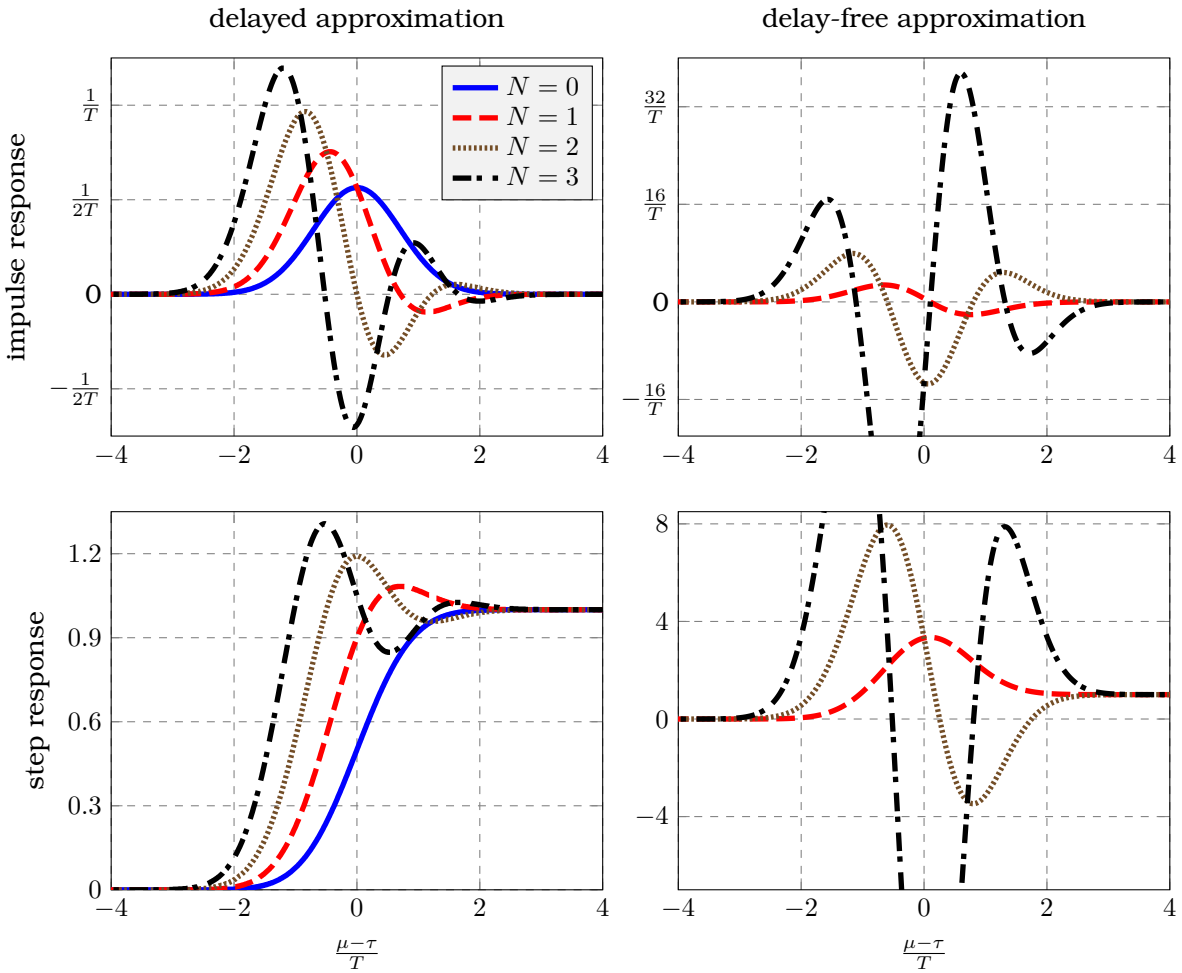


Figure 2.2: Variation of the impulse and step responses given in (2.43) and (2.45), respectively, of a Hermite diff. as a function of the parameter N.

rewritten as

$$g_{N,T,\vartheta,\mu} = \sum_{m=0}^{N} \sum_{l=0}^{\lfloor \frac{m}{2} \rfloor} \frac{m!\,\mathrm{H}_m(\vartheta)\,(-1)^l 2^{m-2l}}{T\,\|\mathrm{H}_m\|_w\, l!(m-2l)!}\,(w \cdot \pi_{m-2l}) \circ \theta_T, \tag{2.43}$$

using #A.1. For $N = 0$ it satisfies

$$g_{0,T,\vartheta,\mu} : \tau \mapsto g_{0,T,\vartheta,\mu}(\tau) = \frac{1}{T\sqrt{\pi}} \mathrm{e}^{-\left(\frac{\mu-\tau}{T}\right)^2}, \tag{2.44}$$

hence, the normal distribution with mean value μ and variance $T^2/2$ is recovered.

The symmetry properties of the Hermite polynomials given in #A.2 yield

$$g_{N,T,-\vartheta,\mu}(\tau+\mu) = g_{N,T,\vartheta,\mu}(-\tau+\mu).$$

Thus, changing the sign of ϑ, which corresponds to mirroring the estimation delay around the point μ, yields the reflection of the diff. across the line μ.

The step response $h_{N,T,\vartheta,\mu}$ can be computed using the properties of the Hermite polynomi-

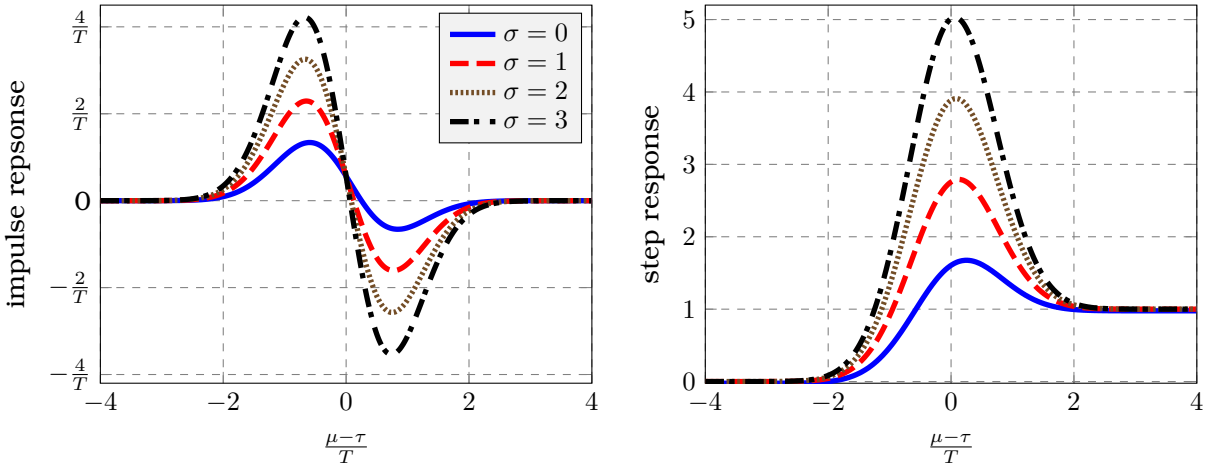


Figure 2.3: Variation of the impulse and step responses of a delay-free Hermite diff. with respect to the parameter $\mu = \sigma\kappa$ and $N = 1$.

als from #A.5 to get

$$\begin{aligned} h_{N,T,\vartheta,\mu}(\tau) &= \int_0^\tau g_{N,T,\vartheta,\mu}(\sigma)\mathrm{d}\sigma \\ &= \frac{1}{2}\left(\operatorname{erf}\left(\frac{\mu}{T}\right) - \operatorname{erf}\left(\frac{\mu-\tau}{T}\right)\right) \\ &\quad - \sum_{m=1}^{N} \frac{\mathrm{H}_m(\vartheta)}{\|\mathrm{H}_m\|_w}\left(w\left(\frac{\mu}{T}\right)\mathrm{H}_{m-1}\left(\frac{\mu}{T}\right) - w\left(\frac{\mu-\tau}{T}\right)\mathrm{H}_{m-1}\left(\frac{\mu-\tau}{T}\right)\right) \end{aligned} \tag{2.45}$$

with $\tau \mapsto \operatorname{erf}(\tau)$ the Error Function defined in #A.15.

The impulse and step responses of Hermite diffs. are shown in Figs. 2.2 and 2.3 for different parameter combinations. The impulse response depicts negative values for $N > 0$, which results in overshoots in the step responses. The magnitude of the overshoot peaks increase with increasing N. It can be seen that the step response in the delay-free case shows significant overshoots.

Frequency-domain properties: The Fourier transform of a Hermite diff. can be computed using the property in #B.11 as

$$\begin{aligned} \mathcal{F}\{g_{N,T,\vartheta,\mu}\}(\omega) &= \frac{1}{T}\sum_{m=0}^{N}\frac{\mathrm{H}_m(\vartheta)}{\|\mathrm{H}_m\|_w}\mathcal{F}\{(w\cdot\mathrm{H}_m)\circ\theta_T\}(\omega) \\ &= \sqrt{\pi}\mathrm{e}^{-\iota\omega\mu}w\left(\frac{T\omega}{2}\right)\sum_{m=0}^{N}\frac{\mathrm{H}_m(\vartheta)}{\|\mathrm{H}_m\|_w}(\iota T\omega)^m, \end{aligned} \tag{2.46}$$

where $\iota^2 = -1$. Since $\omega \mapsto \mathcal{F}\{g_{N,T,\vartheta,\mu}\}(\omega)$ is continuously differentiable it follows that $\mathcal{F}\{h\}$ with h defined in (2.38) also has this property. It is then easy to show that the admissibility condition in (2.39) is satisfied. Thus, the wavelet interpretation given in Section 2.3.5 is valid for this choice of orth. diffs.

The variation of the amplitude spectrum of the diff., i.e., $\omega \mapsto |\mathcal{F}\{g_{N,T,\vartheta,\mu}\}(\omega)|$, as a function of N is given in Figs. 2.4 and 2.6. As for the step response, it can be seen that the amplitude spectrum shows overshoots for $N > 0$. They are significant for the delay-free case and may

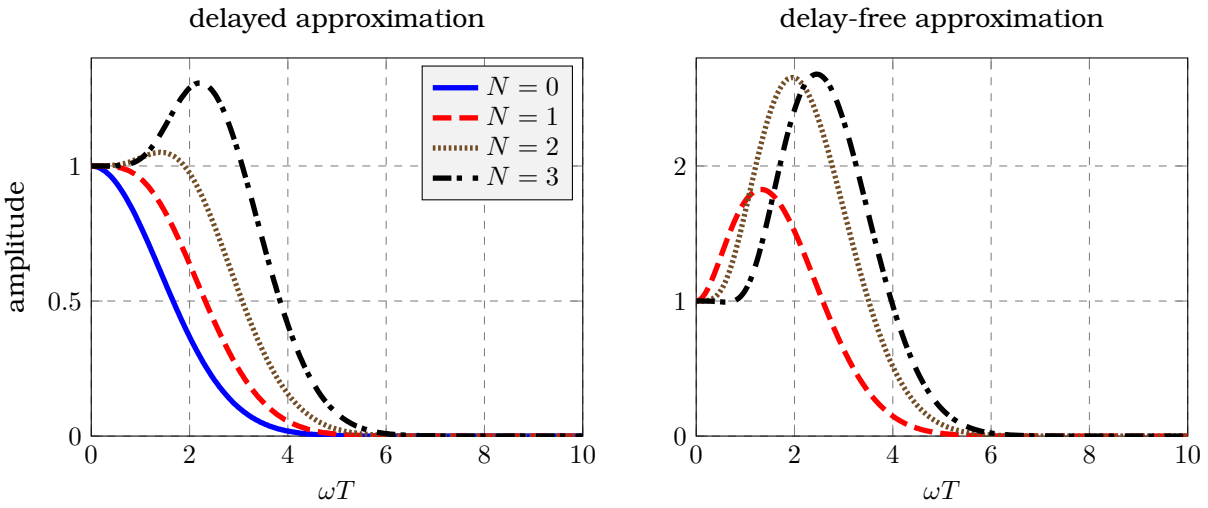


Figure 2.4: Variation of the amplitude spectrum of a Hermite diff. for different values of the parameter N.

yield useless results in practical implementations. The amplitude spectrum also shows a low-pass behaviour: For low frequencies the amplitude spectrum is close to one since

$$\lim_{\omega \to 0} \mathcal{F}\left\{g_{N,T,\vartheta,\mu}\right\}(\omega) = 1.$$

High frequencies are attenuated.

This frequency-domain analysis is used in the next chapter to derive tuning guidelines.

2.4.2 Laguerre differentiators

The Laguerre polynomials are orthogonal over the interval $\mathcal{I}_{w_\alpha} = [0, \infty)$. Their weight function $w_\alpha : t \mapsto w_\alpha(t) = \mathrm{e}^{-t} t^\alpha$, $\alpha > -1$, satisfies $\lim_{\tau \to \infty} w_\alpha(\tau) = 0$, i.e., the effects of w_α vanishes for large arguments. The interval has to be flipped and shifted. A degree of freedom can be incorporated by adding a scaling parameter T. From Table 2.2, the transformation follows as

$$\begin{aligned} \varphi_{t,T} : \mathcal{I}_{w_\alpha} &\to \underline{\mathcal{I}(t)}, \\ \tau &\mapsto \varphi_{t,T}(\tau) = -T\tau + t, \end{aligned} \tag{2.47a}$$

with the inverse

$$\begin{aligned} \varphi_{t,T}^{-1} : \underline{\mathcal{I}(t)} &\to \mathcal{I}_{w_\alpha}, \\ \tau &\mapsto \varphi_{t,T}^{-1}(\tau) = \frac{t - \tau}{T}, \end{aligned} \tag{2.47b}$$

where $\underline{\mathcal{I}(t)} = (-\infty, t]$. The approximation of a function $f : (-\infty, t] \to \mathbb{R}$ using orthogonal Laguerre polynomials, is now discussed. Their properties are summarised in Table 2.1.

The filter interpretation developed in (2.35) of the approximation scheme and the definitions (2.18), (2.21), and (2.28) can be used to derive the delayed approximation $\hat{f}$ of f as

$$\hat{f}(t) = \int_{-\infty}^{t} g_{N,T,\vartheta}^{(\alpha)}(t - \tau) f(\tau) \mathrm{d}\tau \tag{2.48}$$

with

$$g_{N,T,\vartheta}^{(\alpha)}(\tau) = \frac{1}{T} \sum_{n=0}^{N} \frac{\mathrm{L}_n^{(\alpha)}(\vartheta)}{\left\| \mathrm{L}_n^{(\alpha)} \right\|_{w_\alpha}^2} \left(w_\alpha \cdot \mathrm{L}_n^{(\alpha)} \right) \circ \theta_T(\tau),$$

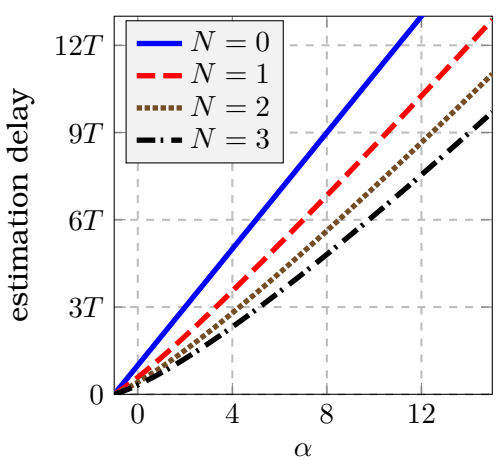


Figure 2.5: Variation of the estimation delay δ_t of Lag. diffs. as a function of the parameters N and α.

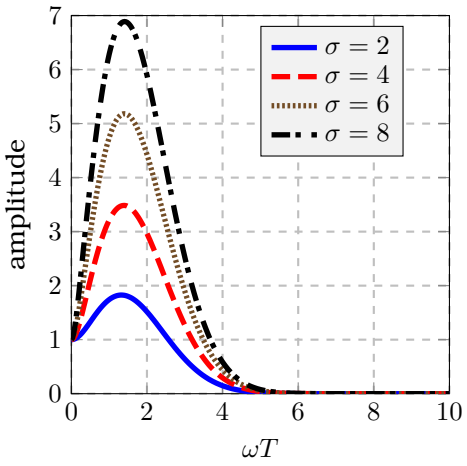


Figure 2.6: Variation of the amplitude spectrum of delay-free Hermite diffs. with respect to the parameter $\mu = \sigma T$ and $N = 1$.

where

$$\left\|\mathrm{L}_n^{(\alpha)}\right\|_{w_\alpha} = \frac{\Gamma(n+\alpha+1)}{n!}, \quad w_\alpha(\tau) = \mathrm{e}^{-\tau}\tau^\alpha, \quad \theta_T(\tau) = \frac{\tau}{T}, \quad \vartheta \in \mathbb{R} \quad \text{and} \quad \alpha > -1.$$

The parameter ϑ influences the estimation delay and the degree of exactness as discussed in Corollaries 2.2 and 2.3. Choosing ϑ as a zero of the Laguerre OP of degree $N+1$ decreases the approximation error by one order as discussed in Corollaries 2.1 and 2.4 and yields a delayed approximation. The estimation delay is given in Table 2.3 and its variation with respect to the parameters N and α is shown in Fig. 2.5. Increasing α increases the delay and in accordance with Theorem 2.1, increasing N decreases the delay. A delay-free approximation is achieved by choosing $\vartheta = 0$.

For the approximation of the n-th order derivative of a sufficiently differentiable functions f it has been shown in Section 2.3.1 that the $n-1$ first derivatives of the kernel $g_{N,T,\vartheta,\mu}$ must vanish at the boundaries of $\mathcal{I}_w$. Recalling the analytical form of the derivatives of the Laguerre polynomials from #A.8 yields

$$\left(g_{N,T,\vartheta}^{(\alpha)}\right)^{(l)} = \frac{1}{T^{l+1}} \sum_{m=0}^{N} \frac{(m+l)!\,\mathrm{L}_m^{(\alpha)}(\vartheta)}{m!\left\|\mathrm{L}_m^{(\alpha)}\right\|_w^2} \left(w_{\alpha-l} \cdot \mathrm{L}_{m+l}^{(\alpha-l)}\right) \circ \theta_T, \tag{2.49}$$

for $\alpha > n-1$ and $l \in \{1, \ldots, n-1\}$, and it is clear that the derivatives of the kernel vanishes.

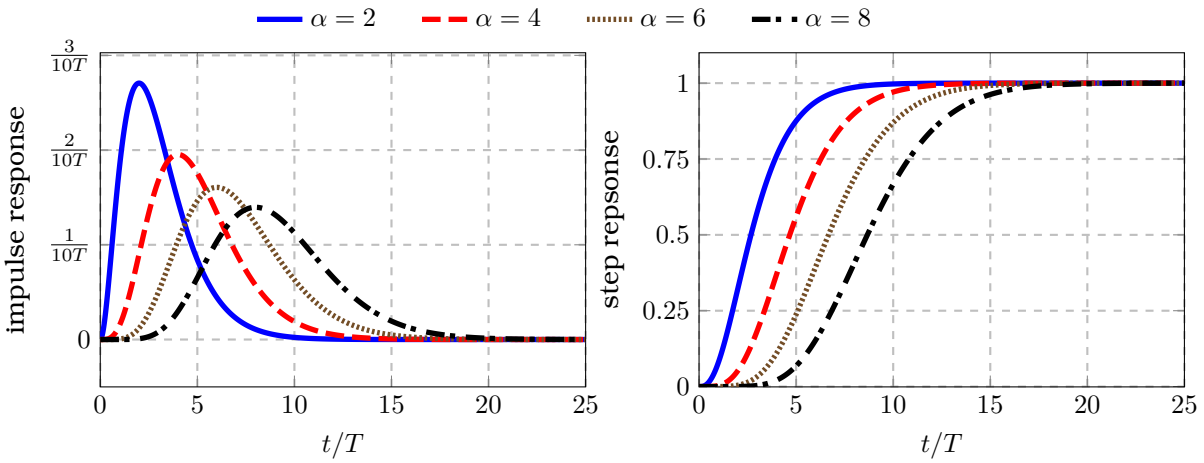


Figure 2.7: Impulse and step responses $g = \bar{g}^{(\alpha)}_{N,T,\vartheta}$ and $h = h^{(\alpha)}_{N,T,\vartheta}$ from (2.53) and (2.55), respectively, of a Lag. diff. with $N = 0$ and various values of α.

Definition 2.7: Laguerre diff.

Let $N, n \in \mathbb{N}$ and $T, \alpha, \vartheta \in \mathbb{R}$, with $T > 0$ and $\alpha > n - 1$. The function

$$g^{(\alpha)}_{N,T,\vartheta} : \tau \mapsto g^{(\alpha)}_{N,T,\vartheta}(\tau) = g^{(\alpha)}_{N,T,\vartheta}(\tau) = \frac{1}{T} \sum_{n=0}^{N} \frac{\mathrm{L}^{(\alpha)}_n(\vartheta)}{\left\| \mathrm{L}^{(\alpha)}_n \right\|^2_{w_\alpha}} \left(w_\alpha \cdot \mathrm{L}^{(\alpha)}_n \right) \circ \theta_T(\tau), \tag{2.50a}$$

with

$$\left\| \mathrm{L}^{(\alpha)}_n \right\|_{w_\alpha} = \frac{\Gamma(n + \alpha + 1)}{n!}, \quad w_\alpha(\tau) = \mathrm{e}^{-\tau} \tau^\alpha, \quad \text{and} \quad \theta_T(\tau) = \frac{\tau}{T}, \tag{2.50b}$$

is called a Laguerre differentiator (Lag. diff.) of order n. Its n-th order derivative is given in (2.49).

Filter characteristics

As for the function approximation based on Hermite polynomials the filter properties of Lag. diffs. are now discussed in the time and frequency domains. The properties of the filter $g^{(\alpha)}_{N,T,\vartheta}$ and the cascade of several filters for the estimation of higher order derivatives are analysed.

Time-domain properties: For $N = 0$ a Lag. diff. $g^{(\alpha)}_{0,T,\vartheta}$ defined in Definition 2.7 can be written as

$$g^{(\alpha)}_{0,T,\vartheta} = \frac{w_\alpha \circ \theta_T}{T\,\Gamma(\alpha + 1)}, \tag{2.51}$$

with Γ the Gamma function defined in #A.13. Recalling the explicit definition of Laguerre polynomials from #A.6, any Lag. diff. $g^{(\alpha)}_{N,T,\vartheta}$ can be rewritten as

$$g^{(\alpha)}_{N,T,\vartheta} = \sum_{n=0}^{N} \sum_{m=0}^{n} l^{(\alpha)}_{n,m,\vartheta} g^{(\alpha+m)}_{0,T,\vartheta} \tag{2.52a}$$

with

$$l^{(\alpha)}_{n,m,\vartheta} = \frac{(-1)^m \Gamma(\alpha + m + 1) \mathrm{L}^{(\alpha)}_n(\vartheta)}{m! \left\| \mathrm{L}^{(\alpha)}_n \right\|^2_{w_\alpha}} \binom{n + \alpha}{n - m}. \tag{2.52b}$$

Observing that the latter sum has a triangular structure yields

$$g_{N,T,\vartheta}^{(\alpha)} = \sum_{n=0}^{N} \bar{l}_{n,\vartheta}^{(\alpha)} g_{0,T,\vartheta}^{(\alpha+n)}, \quad \text{with} \quad \bar{l}_{n,\vartheta}^{(\alpha)} = \sum_{m=n}^{N} l_{m,n,\vartheta}^{(\alpha)}.$$

The impulse response of a Lag. diff. $g_{N,T,\vartheta}^{(\alpha)}$ is

$$\bar{g}_{N,T,\vartheta}^{(\alpha)}(\tau) = \begin{cases} g_{N,T,\vartheta}^{(\alpha)}(\tau), & \text{for} \quad \tau \geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.53}$$

The special parametrisation $N = 0$ yields

$$\bar{g}_{0,T,\vartheta}^{(\alpha)}(\tau) = \begin{cases} \frac{1}{T\Gamma(\alpha+1)} \mathrm{e}^{-\frac{\tau}{T}} \left(\frac{\tau}{T}\right)^{\alpha}, & \text{for} \quad \tau \geqslant 0, \\ 0, & \text{otherwise,} \end{cases}$$

and corresponds to the probability density function in the shape-scale parametrisation of a Gamma distribution (see, e.g., Papoulis (1984)) with mean and variance

$$\eta = (\alpha + 1)T, \tag{2.54a}$$

$$\sigma^2 = (\alpha + 1)T^2, \tag{2.54b}$$

respectively. Furthermore, it also follows from Hegland and Anderssen (1998) that for this choice of parameters the diff. is a mollifier and derivative estimation is a regularisation strategy.

The step response $h_{N,T,\vartheta}^{(\alpha)}$ of a Lag. diff. $g_{N,T,\vartheta}^{(\alpha)}$ is

$$\begin{aligned} h_{N,T,\vartheta}^{(\alpha)}(t) &= \int_0^t g_{N,T,\vartheta}^{(\alpha)}(\tau) \mathrm{d}\tau \\ &= \frac{1}{\Gamma(\alpha+1)} \gamma\left(\alpha + 1, \frac{t}{T}\right) + \sum_{n=1}^{N} \frac{\mathrm{L}_n^{(\alpha)}(\vartheta)}{\left\| \mathrm{L}_n^{(\alpha)} \right\|_{w_\alpha}^2} \int_0^{\frac{t}{T}} w_\alpha(\tau) \mathrm{L}_n^{(\alpha)}(\tau) \mathrm{d}\tau, \end{aligned}$$

with γ the lower incomplete Gamma function defined in #A.14. With relation #A.9 the step response can be written as

$$h_{N,T,\vartheta}^{(\alpha)}(t) = \frac{1}{\Gamma(\alpha+1)} \gamma\left(\alpha + 1, \frac{t}{T}\right) + \sum_{n=1}^{N} \frac{\mathrm{L}_n^{(\alpha)}(\vartheta)}{\left\| \mathrm{L}_n^{(\alpha)} \right\|_{w_\alpha}^2} \frac{1}{n} w_{\alpha+1}(t) \mathrm{L}_{n-1}^{(\alpha+1)}(t). \tag{2.55}$$

Figs. 2.7 to 2.9 show the impulse and step responses of several diffs. As for Hermite diffs. significant overshoots appear in the step responses when $\vartheta = 0$, i.e, for the delay-free case. These overshoots decrease with increasing α. Contrarily, the overshoots are negligible when $\vartheta = l_{N+1}^{(\alpha)}$, i.e., when the approximation is delayed. While for $\alpha = 0$ the maximum of the impulse response is at the beginning of the interval, for increasing α it moves to the right. Thus, for $\alpha = 0$ the most recent values of f are multiplied with the strongest weights, which is not true any more when α is increased. This corresponds to the increase of the estimation delay for increasing α as shown in Fig. 2.5.

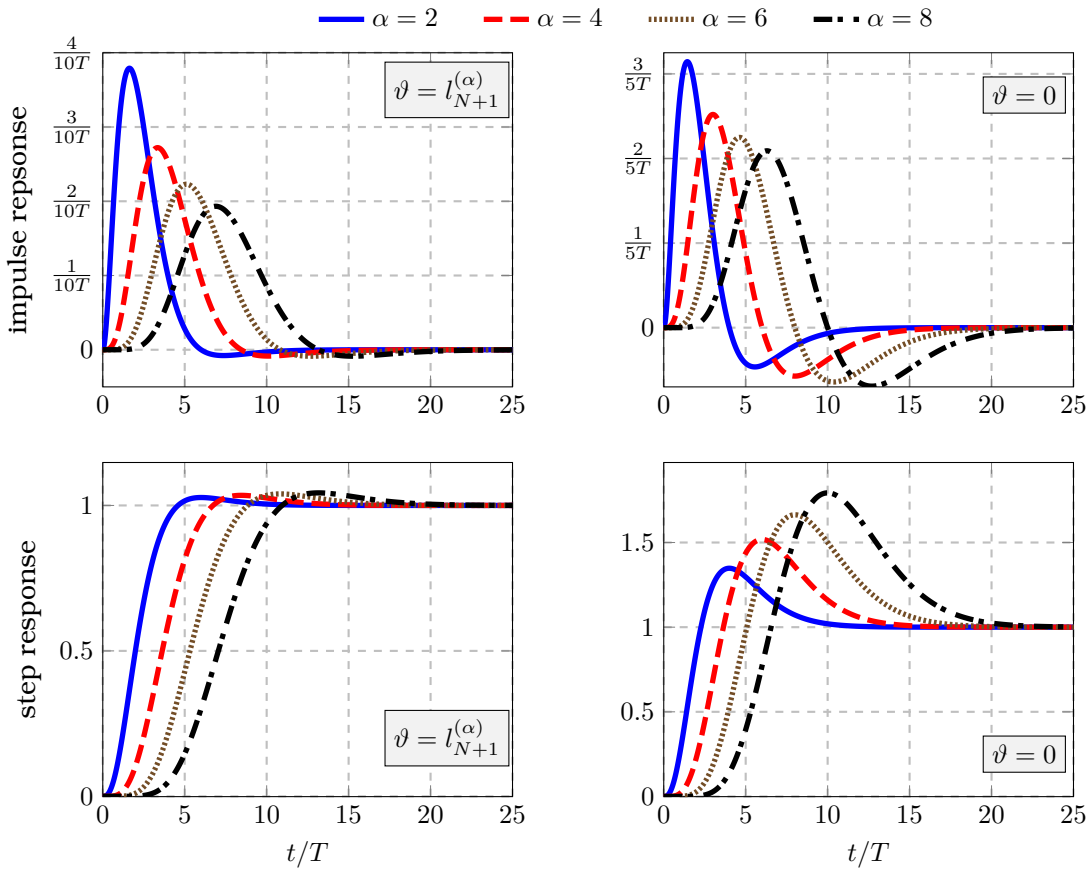


Figure 2.8: Impulse and step responses of a Lag. diff. for $N = 1$ and various values of α. Differentiators with delay are shown on the left and delay-free ones on the right.

Cascade of diffs. Let $\bar{g}_i = \bar{g}_{0,T_i,\vartheta}^{(\alpha_i)}$ and recall that it is a density function with mean value η_i and variance σ_i^2 given in (2.54). Consider a cascade of $p \in \mathbb{N}$ diffs. The impulse response G_p of this cascade can be written as the convolution product of the impulse responses $\bar{g}_i$ of the diffs. From (Papoulis, 1962, Sec. 11.2) it follows that the first and second moments of G_p are

$$\eta(p) = \sum_{n=1}^{p} \eta_i \quad \text{and} \quad \sigma^2(p) = \sum_{n=0}^{p} \sigma_i^2. \tag{2.56}$$

Since the third moments

$$m_{3,i} = \int_{-\infty}^{\infty} \tau^3 \bar{g}_i(\tau) \mathrm{d}\tau = T^3(\alpha+1)(\alpha+2)(\alpha+3)$$

of $\bar{g}_i$, $i \in \{1, \ldots, p\}$, are finite and $\lim_{p\to\infty} \sigma^2(p) = \infty$, by using the central-limit theorem (see, e.g., (Papoulis, 1962, Sec. 11.2)) it can be concluded that G_p tends to a Gaussian function, i.e.,

$$G_p(t) \to \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(t-\eta(p))^2/(2\sigma^2(p))} \quad \text{for} \quad p \to \infty.$$

Define the duration D_t of a function $f : t \mapsto f(t)$ by $D_t\{f\} = \int_{-\infty}^{\infty} t^2 |f(t)| \mathrm{d}t$, the duration D_ω of its Fourier transform $\mathcal{F}$ by $D_\omega\{f\} = \int_{-\infty}^{\infty} \omega^2 \, |\mathcal{F}(\omega)| \, \mathrm{d}\omega$, and its energy by $\mathcal{E}\{f\} = \int_{-\infty}^{\infty} |f(t)|^2 \mathrm{d}t$. The uncertainty principle (see (Papoulis, 1962, Sec. 4.4)) states that if f satisfies $\lim_{t\pm\infty} \sqrt{t} f(t) = 0$, then

$$\frac{D_t\{f\} D_\omega\{f\}}{\mathcal{E}\{f\}} \geqslant \sqrt{\frac{2}{\pi}},$$

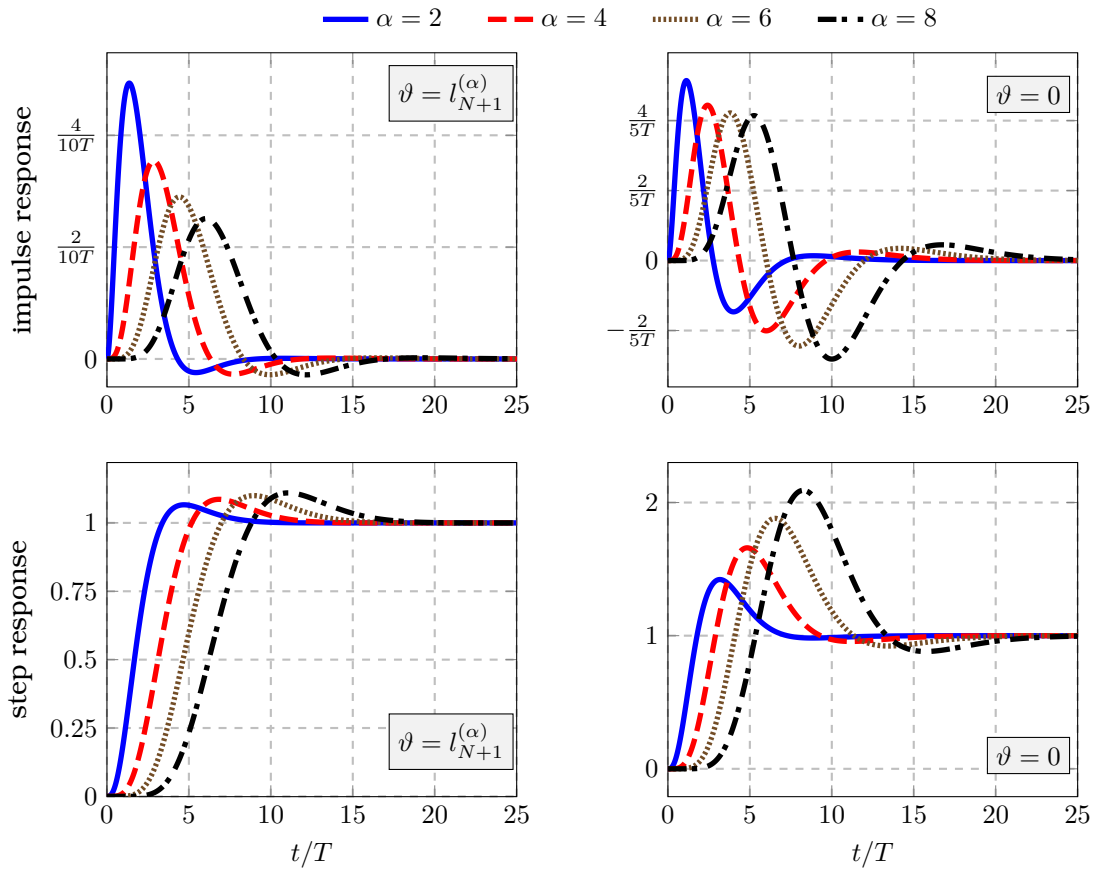


Figure 2.9: Impulse and step responses of a Lag. diff. for $N = 2$ and various values of α. Diffs. with delay are shown on the left and delay-free ones on the right.

and the equality holds only for Gaussian $f(t) = \sqrt{\frac{\beta}{\pi}}\mathrm{e}^{-\beta t^2}$, $\beta > 0$. Thus, the cascade G_p converges to a compromise between the width of the impulse response, i.e. a clear concentration of values to be filtered and thus a high estimation accuracy, and the width of the frequency response, which corresponds to an effective low-pass effect.

Corollary 2.6: Optimality of cascade of Laguerre diffs.

Consider a cascade of $p \in \mathbb{N}$ diffs. The impulse response G_p of this cascade converges to an optimal compromise in the sense of the uncertainty principle between the resolution in the time and the frequency domain for $p \to \infty$.

It follows that the cascade of first-order Laguerre diffs. for the estimation of higher order derivatives can yield better results. Fig. 2.10 shows the impulse response of a cascade of Lag. diffs. It can be seen that the responses converge quickly to a Gaussian function.

To parametrise a diff. $g_{0,\vartheta,T}^{(\alpha)}$ with desired mean value η and variance σ^2 it is sufficient to invert (2.54a) and (2.54b) to get

$$T_\mathrm{d} = \sigma^2/\eta \quad \text{and} \quad \alpha_\mathrm{d} = \frac{\eta^2}{\sigma^2} - 1. \tag{2.57}$$

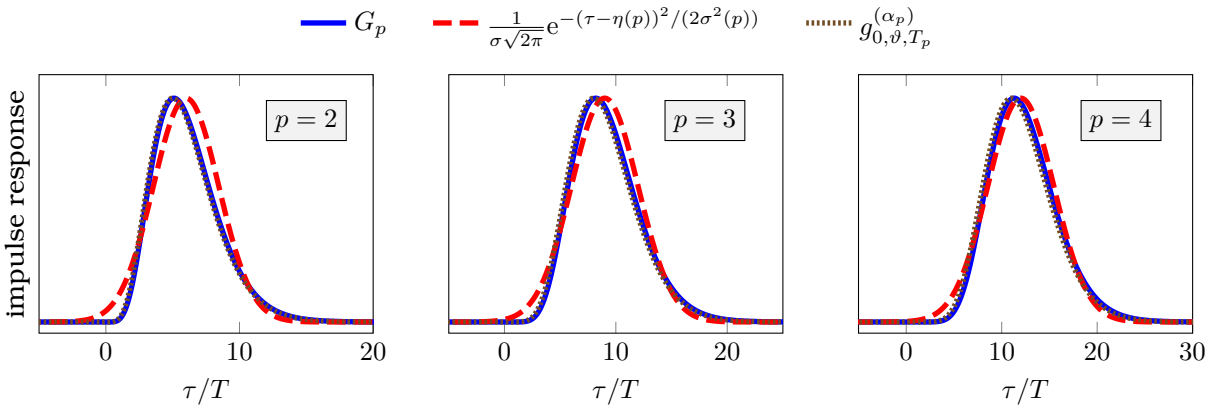


Figure 2.10: Impulse responses of G_p, the cascade of p Lag. diffs. with $\alpha = 4$, a diff. $g^{(\alpha_p)}_{0,\vartheta,T_p}$ with α_p and T_p from (2.58), and a Gaussian for different values of p. The mean $\eta(p)$ and the variance $\sigma(p)$ are defined in (2.56).

Thus, using (2.54) and (2.57) a Lag. diff. $g^{(\alpha_p)}_{0,\vartheta,T_p}$ can be parametrised by

$$T_p = T \quad \text{and} \quad \alpha_p = p(\alpha+1) - 1 \tag{2.58}$$

to get the properties of the cascade G_p. Fig. 2.10 shows the impulse response of $g^{(\alpha_p)}_{0,\vartheta,T_p}$ for different values of p. It can be seen that, as expected, $g^{(\alpha_p)}_{0,\vartheta,T_p}$ converges to a Gaussian function. Thus, for $\alpha \to \infty$ and $N = 0$ the Hermite diffs. are recovered from the Lag. diffs.

Frequency-domain properties: Using the Fourier transform of the function

$$f : t \mapsto f(t) = \begin{cases} \mathrm{L}_n^{(\alpha)}(t)\mathrm{e}^{-t}t^{\alpha}, & \text{for} \quad t \geqslant 0 \\ 0, & \text{otherwise,} \end{cases}$$

given in #B.13, the Fourier transform $\omega \mapsto \mathcal{F}\left\{\bar{g}^{(\alpha)}_{N,T,\vartheta}\right\}$ of $\bar{g}^{(\alpha)}_{N,T,\vartheta}$ can be computed as

$$\mathcal{F}\left\{\bar{g}^{(\alpha)}_{N,T,\vartheta}\right\}(\omega) = \sum_{n=0}^{N} \mathrm{L}_n^{(\alpha)}(\vartheta)\frac{(\iota\omega T)^n}{(\iota\omega T+1)^{(\alpha+n+1)}}. \tag{2.59}$$

Since $\omega \mapsto \mathcal{F}\left\{\bar{g}^{(\alpha)}_{N,T,\vartheta}\right\}(\omega)$ is continuously differentiable it follows that $\mathcal{F}\{h\}$ with h defined in (2.38) also has this property. It is then easy to show that the admissibility condition in (2.39) is satisfied. Thus, the wavelet interpretation given in Section 2.3.5 is valid when Laguerre polynomials are used.

Fig. 2.11 shows different amplitude spectra of Lag. diffs. with and without delay. It can be seen that in the delay-free case significant overshoots appear. Delay-free diffs. amplify the low-frequency content of the signals by more than 100% (e.g, by 280% for $N = 2$ and $\alpha = 4$), which might yield very bad approximations and may not be tolerable in practice.

Annihilators

The algebraic interpretation of numerical differentiation discussed in Section 2.3.2 is now specified for Lag. diff. and an annihilator of the form (2.33) is derived. As in Section 2.3.2, an analytic function

$$f : \mathbb{R}_+ \to \mathbb{R} \tag{2.60}$$

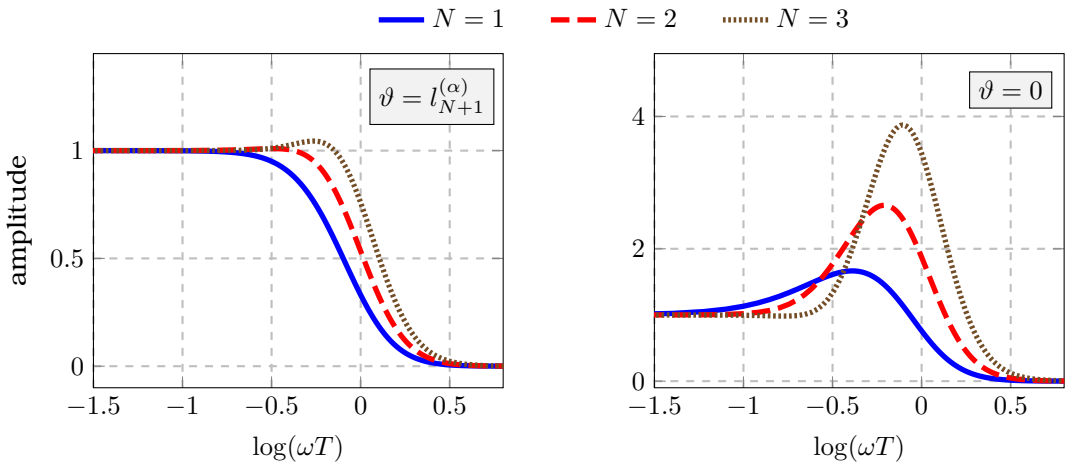


Figure 2.11: Amplitude spectra of Lag. diffs. for $\alpha = 4$ and different values of N. Differentiators with delay are shown on the left and delay-free ones on the right.

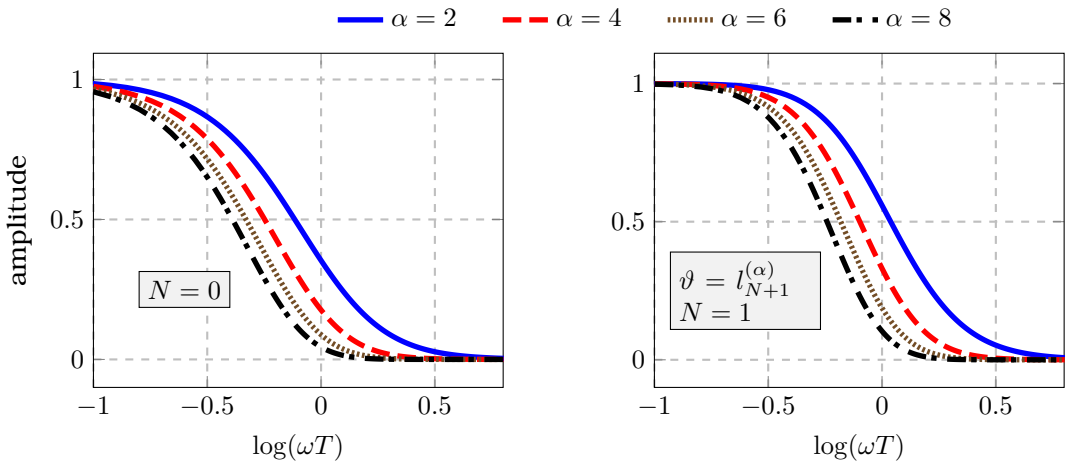


Figure 2.12: Amplitude spectra of Lag. diffs. for different values of α and N. All diffs. exhibit a delay in the approximation.

is considered. In the operational domain, the truncated Taylor series expansion (2.31) corresponds to

$$\mathcal{L}\{f_{n^\star}\}(\mathrm{s}) = \sum_{m=0}^{n^\star} \frac{1}{\mathrm{s}^{1+m}} f^{(m)}(0). \tag{2.61}$$

Recall for convenience that all derivatives, that are not of interest are now considered as undesired perturbations to be annihilated. Assume for the beginning that $n^\star = n$ and consider the differential operator

$$\mathcal{D}_{\mathrm{L},0} = \rho_0(\mathrm{s}) \left(\frac{\mathrm{d}}{\mathrm{ds}}\right)^{n+\alpha_0} \mathrm{s}^n$$

with $\rho_0(\mathrm{s}) = 1/(\mathrm{s}-1)$ and $\alpha_0 \in \mathbb{N}$. In the sequel, let $\bar{\alpha}_0 = n + \alpha_0$. Applying the operator $\mathcal{D}_{\mathrm{L},0}$ to (2.61) yields

$$\begin{aligned} \mathcal{D}_{\mathrm{L},0}\mathcal{L}\{f_N\}(\mathrm{s}) &= \rho_0(\mathrm{s}) \left(\frac{\mathrm{d}}{\mathrm{ds}}\right)^{\bar{\alpha}_0} \left(\sum_{m=0}^{n} \mathrm{s}^{n-m-1} f^{(m)}(0)\right) \\ &= (-1)^{\bar{\alpha}_0} \bar{\alpha}_0! \frac{\rho_0(\mathrm{s})}{\mathrm{s}^{\bar{\alpha}_0+1}} f^{(n)}(0). \end{aligned} \tag{2.62}$$

Thus, the differential operator $\mathcal{D}_{\mathrm{L},0}$ and the function $\mathrm{s} \mapsto (-1)^{\bar{\alpha}_0}\bar{\alpha}_0!\rho_0(\mathrm{s})/\mathrm{s}^{\bar{\alpha}_0+1}$ are associated.

Writing (2.62) in the time domain and replacing f_N by f, as in Section 2.3.2, yields

$$\begin{aligned}\int_0^t \mathrm{e}^{-\tau}\tau^{\bar{\alpha}_0} f^{(n)}(\tau)\mathrm{d}\tau &= (-1)^{\bar{\alpha}_0} f^{(n)}(0)\int_0^t \mathrm{e}^{-\tau}\tau^{\bar{\alpha}_0}\mathrm{d}\tau,\\ &= (-1)^{\bar{\alpha}_0}\gamma\left(\bar{\alpha}_0+1,t\right) f^{(n)}(0).\end{aligned}$$

Thus, an approximation of the n-th order derivative is

$$\hat{f}^{(n)}(0) = \frac{(-1)^{\bar{\alpha}_0}}{\gamma\left(\bar{\alpha}_0+1,t\right)}\int_0^t \mathrm{e}^{-\tau}\tau^{\bar{\alpha}_0} f^{(n)}(\tau)\mathrm{d}\tau.$$

Recalling from #A.14 that $\lim_{t\to\infty}\gamma\left(\bar{\alpha}_0+1,t\right) = \Gamma(\bar{\alpha}_0+1) = \bar{\alpha}_0!$ and assuming that the integral on the right side of the latter equation exists[16], the approximation of the desired derivative after an infinite time is

$$\hat{f}^{(n)}_{\infty}(0) = \frac{1}{(-1)^{\bar{\alpha}_0}\bar{\alpha}_0!}\int_0^{\infty} \mathrm{e}^{-\tau}\tau^{\bar{\alpha}_0} f^{(n)}(\tau)\mathrm{d}\tau.$$

To obtain a causal approximation of the derivate at time t it is now sufficient to replace $f^{(n)}(\tau)$ in the latter equation by $(-1)^{\bar{\alpha}_0} f^{(n)}\left(t-\tau T\right)\Theta\left(t-\tau T\right)$, where the parameter $T>0$ can be used to dilate or contract the interval $[0,\infty)$. The obtained approximated derivative is

$$\begin{aligned}\hat{f}^{(n)}(t) &= \frac{1}{\bar{\alpha}_0!}\int_0^{\infty} \mathrm{e}^{-\tau}\tau^{\bar{\alpha}_0} f^{(n)}\left(t-\tau T\right)\Theta\left(t-\tau T\right)\mathrm{d}\tau\\ &= \int_0^t g^{(\alpha_0+n)}_{0,T,\vartheta}(t-\tau) f^{(n)}(\tau)\mathrm{d}\tau,\end{aligned} \tag{2.63}$$

with $g^{(\alpha_0+n)}_{0,T,\vartheta}$ defined in (2.51), and corresponds to the approximated (2.48) derived using the approximation theoretic approach. Thus, the Lag. diff. defined in Definition 2.7 with $\alpha\in\mathbb{N}$ and $N=0$ can also be derived using differential-algebraic manipulations of truncated Taylor series when the function of which the derivative is to be approximated is analytic.

The case of a Taylor expansion of order $n^\star > n$ is considered by

1. extending the differential operator in (2.60) to also annihilate the derivatives of orders greater than n, and by
2. exploiting the property (2.52) stating that any Lag. diff. with $N>0$ can be written as a linear combination of minimal ones.

For this purpose, start with the differential operator

$$\mathcal{D}_{\mathrm{L},l} = \rho_0(\mathrm{s})\left(\frac{\mathrm{d}}{\mathrm{ds}}\right)^{n+l}\frac{1}{\mathrm{s}}\left(\frac{\mathrm{d}}{\mathrm{ds}}\right)^{n^\star-n}\mathrm{s}^{n^\star+1}$$

that annihilates all derivatives that are not of interest. This operator is associated to the function $\rho_l : \mathrm{s}\mapsto\rho_l(\mathrm{s}) = (n^\star-n)!(n+l)!(-1)^{n+l}\rho_0(\mathrm{s})/\mathrm{s}^{n+l+1}$. Consider now the differential operator

$$\mathcal{D}_{\mathrm{L}} = \sum_{n=0}^{N}\sum_{m=0}^{n}\frac{l^{(\alpha)}_{n,m,\vartheta}}{(n^\star-n)!}\mathcal{D}_{\mathrm{L},\alpha+m+1-n},\quad \alpha\in\mathbb{N}_{>n-1},$$

[16] Recall that this corresponds to the assumption on $f^{(n)}$ used for the derivation of the diffs. using the approximation-theoretic approach in Section 2.1.2.

with $l^{(\alpha)}_{n,m,\vartheta}$ defined in (2.52b), motivated by the linear combination of minimal diffs. in (2.52). This differential operator is associated to the function

$$\rho : \mathrm{s} \mapsto \rho(\mathrm{s}) = \sum_{n=0}^{N} \sum_{m=0}^{n} \frac{l^{(\alpha)}_{n,m,\vartheta}}{(n^\star - n)!} \rho_l(\mathrm{s}).$$

Exploiting the shifting of the origin as in (2.63), a dilatation parameter T, and the property (2.52) yields the approximated derivative

$$\hat{f}^{(n)}(t) = \int_0^t g^{(\alpha)}_{N,T,\vartheta}(t-\tau) f^{(n)}(\tau) \mathrm{d}\tau.$$

Thus, with a suitable combination of differential operators, Lag. diff. with $\alpha \in \mathbb{N}$ can also be derived using differential-algebraic manipulations on truncated Taylor series assuming that the function of which the derivative is to be approximated is analytic.

2.4.3 Jacobi differentiators

Recall from Table 2.1 that Jacobi polynomials are orthogonal on the interval $\mathcal{I}_{w^{(\alpha,\beta)}} = (-1, 1)$. The time transformation required for the causal approximation of functions follows from Table 2.2 as

$$\begin{aligned} \varphi_{t,T} : \mathcal{I}_w &\to \underline{\mathcal{I}(t)}, \\ \tau &\mapsto \varphi_{t,T}(\tau) = t + \frac{\tau - 1}{2} T, \end{aligned} \tag{2.64a}$$

and its inverse is

$$\begin{aligned} \varphi^{-1}_{t,T} : \underline{\mathcal{I}(t)} &\to \mathcal{I}_w, \\ \tau &\mapsto \varphi^{-1}_{t,T}(\tau) = 1 + 2\frac{\tau - t}{T}, \end{aligned} \tag{2.64b}$$

where $\mathcal{I}_w = [-1, 1]$ and $\underline{\mathcal{I}(t)} = [t - T, t]$.

The results from (2.18), (2.21), (2.28), and (2.35) can be used to derive the approximation $\hat{f}$ of f as

$$\hat{f}(t) = \int_0^T g^{(\alpha,\beta)}_{N,T,\vartheta}(\tau)\, f(t-\tau) \mathrm{d}\tau \tag{2.65a}$$

with

$$g^{(\alpha,\beta)}_{N,T,\vartheta}(\tau) = \frac{2}{T} \sum_{m=0}^{N} \frac{\mathrm{P}^{(\alpha,\beta)}_m(\vartheta)}{\left\| \mathrm{P}^{(\alpha,\beta)}_m \right\|^2} \left(\left(w^{(\alpha,\beta)} \cdot \mathrm{P}^{(\alpha,\beta)}_m \right) \circ \theta_T \right)(\tau), \tag{2.65b}$$

where

$$\left\| \mathrm{P}^{(\alpha,\beta)}_m \right\|^2 = \frac{2^{\alpha+\beta+1} \Gamma(m+\alpha+1)\Gamma(m+\beta+1)}{m!(2m+\alpha+\beta+1)\Gamma(m+\alpha+\beta+1)}, \quad w^{(\alpha,\beta)}(\tau) = (1-\tau)^\alpha (1+\tau)^\beta,$$

and $\theta_T(\tau) = 1 - 2\tau/T$. The parameter ϑ influences the estimation delay and the degree of exactness as discussed in Corollaries 2.2 and 2.3. Choosing ϑ as a zero of the Jacobi OPs of degree $N+1$ decreases the approximation error by one order as discussed in the Corollaries 2.1 and 2.4 and yields a delayed approximation. The estimation delay is given in Table 2.3 and its variation with respect to the parameters α, β, and N is shown in Fig. 2.13. It can be seen that increasing N and β decreases the delay. Contrarily, increasing α increases the delay. These numerical observations can be proved using the properties of OPs in Szegö (1939) as discussed in Kiltz (2017); Mboup (2009); Othmane, Rudolph, and Mounier (2021a), for example. A delay-free approximation is achieved by choosing $\vartheta = 1$.

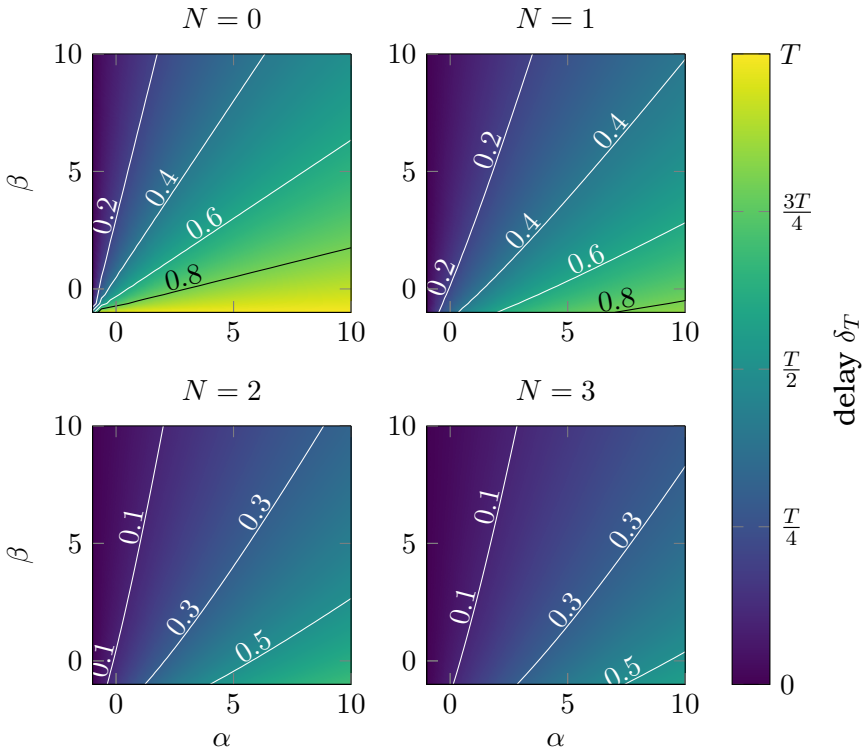


Figure 2.13: Variation of the delay of Jac. diffs. with respect to the parameters α, β, and N.

For the approximation of the n-th order derivatives of a sufficiently differentiable function f it has been discussed in Section 2.3.1 that the $n-1$ first derivatives of $g^{(\alpha,\beta)}_{N,T,\vartheta}$ must vanish at the boundaries of $[t-T,t]$. Recalling the properties of Jacobi polynomials in #A.12 yields

$$\left(g^{(\alpha,\beta)}_{N,T,\vartheta}\right)^{(l)} = \frac{2^{2l+1}}{T^{l+1}} \sum_{m=0}^{N} \frac{(m+l)!\mathrm{P}^{(\alpha,\beta)}_m(\vartheta)}{m!\left\|\mathrm{P}^{(\alpha,\beta)}_m\right\|^2} \left(\mathrm{P}^{(\alpha-l,\beta-l)}_{m+l} \cdot w^{(\alpha-l,\beta-l)}\right) \circ \theta_T \tag{2.66}$$

for $\alpha,\beta > n-1$ and $l \in \{0,\dots,n-1\}$. It is thus clear that for $\alpha,\beta > n-1$ the derivatives always vanish at the boundaries.

Definition 2.8: Jacobi diff.

Let $N,n \in \mathbb{N}$ and $\alpha,\beta,\vartheta \in \mathbb{R}$, with $T > 0$ and $\alpha,\beta > n-1$. The function

$$g^{(\alpha,\beta)}_{N,T,\vartheta} : \tau \mapsto g^{(\alpha,\beta)}_{N,T,\vartheta}(\tau) = \frac{2}{T} \sum_{m=0}^{N} \frac{\mathrm{P}^{(\alpha,\beta)}_m(\vartheta)}{\left\|\mathrm{P}^{(\alpha,\beta)}_m\right\|^2} \left(w^{(\alpha,\beta)} \cdot \mathrm{P}^{(\alpha,\beta)}_m\right) \circ (\theta_T(\tau)),$$

with

$$\left\|\mathrm{P}^{(\alpha,\beta)}_m\right\|^2 = \frac{2^{\alpha+\beta+1}\Gamma(m+\alpha+1)\Gamma(m+\beta+1)}{m!(2m+\alpha+\beta+1)\Gamma(m+\alpha+\beta+1)}, \quad w^{(\alpha,\beta)}(\tau) = (1-\tau)^\alpha(1+\tau)^\beta,$$

and $\theta_T(\tau) = 1 - 2\tau/T$, is called a Jacobi diff. of order n. Its n-th order derivative is given in (2.66).

From this approximation theoretic viewpoint, the estimation process can be explained as

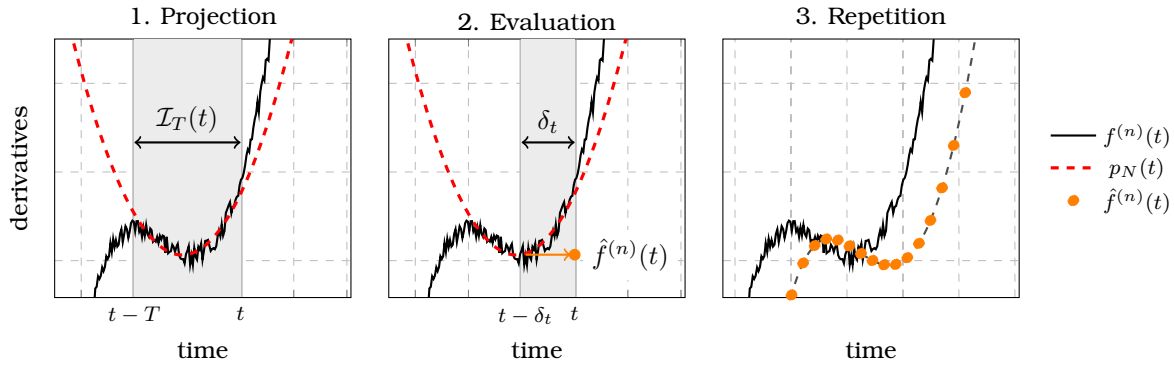


Figure 2.14: Three-step process of the estimation of the n-th order derivative $f^{(n)}$ of a function $f : t \mapsto f(t)$ using Jac. diffs. as given in Othmane et al. (2022).

in Othmane et al. (2022) by the following three steps[17] illustrated in Fig. 2.14:

1. Projection: At time t, the n-th order derivative $f^{(n)}$ on the interval $\mathcal{I}_T(t)$ is projected to the space of polynomials of degree N. This yields the polynomial p_N depicted in the left and middle part of Fig. 2.14.

2. Evaluation: The polynomial p_N is evaluated at $t - \delta_t$, which gives an approximated derivative $\hat{y}^{(n)}(t) = p_N(t - \delta_t)$ for the derivative $y^{(n)}(t)$ as depicted in the central part of Fig. 2.14. Choosing the delay to be a root of the OP increases the approximation order by 1 with a precisely known delay. Alternatively, a delay-free estimation or even a prediction of the future derivative might be selected, at the cost of a reduced accuracy.

3. Repetition: The first two steps are repeated at each time instant t_i while keeping the parameters of the diffs. constant, i.e., evaluating (2.35) at every discrete time instant t_i. This yields the approximated derivative $\hat{f}^{(n)}$ depicted in the right part of Fig. 2.14, where the dots show the evaluation of $\hat{f}^{(n)}$ at discrete time instants.

Filter characteristics

As for the previous diffs. the filter properties of Jac. diffs. are now discussed in the time and frequency domains. The properties of the filter $g^{(\alpha,\beta)}_{N,T,\vartheta}$ and the cascade of several filters for the estimation of higher order derivatives are analysed. The discussions of this section stem almost entirely from Othmane et al. (2022). The reader is referred to the latter reference for a more detailed discussion on the effects of the parameters on the estimation errors.

Time-domain properties: From (2.65) it is clear that the diffs. act as linear finite-impulse response (FIR) filters on the sought derivative. This motivates the analysis of their impulse and step responses.

The closed form of the impulse response $g^{(\alpha,\beta)}_{N,T,\vartheta}$ is given in (2.67), which simplifies to

$$g^{(\alpha,\beta)}_{0,T,\vartheta} = \frac{w^{(\alpha,\beta)} \circ \theta_T}{2^{\alpha+\beta} T \mathrm{B}(\alpha+1, \beta+1)}$$

[17] This time-domain interpretation of the numerical differentiation using Jacobi OPs has first been described in Kiltz (2017); Kiltz and Rudolph (2013).

for $N = 0$. In the latter equation, B is the Beta function defined in #A.16.

Using the explicit definition of Jacobi polynomials given in #A.11, any $g^{(\alpha,\beta)}_{N,T,\vartheta}$ can be written as a linear combination of minimal Jac. diffs. as

$$\begin{aligned} g^{(\alpha,\beta)}_{N,T,\vartheta} &= \sum_{m=0}^{N}\sum_{l=0}^{m}\binom{m+\alpha}{l}\binom{m+\beta}{m-l}\frac{(-1)^{m-l}\mathrm{P}^{(\alpha,\beta)}_m(\vartheta)}{2^{m-1}T\left\|\mathrm{P}^{(\alpha,\beta)}_m\right\|^2}w^{(\alpha+m-l,\beta+l)}\circ\theta_T \\ &= \sum_{m=0}^{N}\sum_{l=0}^{m}\binom{m+\alpha}{l}\binom{m+\beta}{m-l}\frac{(-1)^{m-l}\mathrm{P}^{(\alpha,\beta)}_m(\vartheta)}{2^{m-1}\left\|\mathrm{P}^{(\alpha,\beta)}_m\right\|^2}\mathrm{B}^{(\alpha,\beta)}_{m,l}g^{(\alpha+m-l,\beta+l)}_{0,T,\vartheta} \end{aligned} \tag{2.68}$$

with $\mathrm{B}^{(\alpha,\beta)}_{m,l} = 2^{\alpha+\beta+m}\mathrm{B}(\alpha+m-l+1,\beta+l+1)$.

The step response has been derived in Kiltz (2017) and reads

$$h^{(\alpha,\beta)}_{N,T,\vartheta}(t) = \begin{cases} 0, & \text{for} \quad t<0 \\ I_{\frac{t}{T}}(\bar\alpha,\bar\beta) + \sum_{m=0}^{N}\frac{\mathrm{P}^{(\alpha,\beta)}_m(\vartheta)}{\left\|\mathrm{P}^{(\alpha,\beta)}_m\right\|^2}\left(w^{(\bar\alpha,\bar\beta)}\cdot\mathrm{P}^{(\bar\alpha,\bar\beta)}_{m-1}\right)\circ(\theta_T(\tau))\,, & \text{for} \quad t\in[0,T] \\ 1. & \text{otherwise} \end{cases}$$

with $\bar\alpha = \alpha+1$, $\bar\beta = 1+\beta$, and I_τ the regularised incomplete Beta function defined in #A.18. Figs. 2.15 to 2.17 depict the responses for different parameters. For $N > 0$ the impulse response can become negative and the step response may show an overshoot. This overshoot increases with increasing α and β. As for the previous diffs., the overshoots of a delay-free parametrisation are significant (compare the Figs. 2.16 and 2.17). The variation of the delay as a function of α and β observed in Fig. 2.13 can again be seen in the variation of the impulse response. Increasing α shifts the maximum towards T which increases the approximation delay. Contrarily, increasing β shifts the maximum towards 0 and thus decreases the delay.

For $\alpha = \beta$ and $N = 0$ both are symmetric around the axis $t = T$ or the point $(T/2, 1/2)$, respectively. Recall from Othmane et al. (2022) that this parametrisation corresponds to the Legendre filters. More generally it can be verified that

$$g^{(\beta,\alpha)}_{N,T,\vartheta}\left(t+\frac{T}{2}\right) = g^{(\alpha,\beta)}_{N,T,-\vartheta}\left(\frac{T}{2}-t\right). \tag{2.69}$$

Cascading differentiators: In (Kiltz, 2017, Sec. 3.2.2) it has been proven that a cascade of an infinite number of minimal diffs. converges to a Gaussian kernel G. Let $\bar g_p(t) = g_1(t)g_2(t)\cdots g_p(t)$, where $g_i(t) = g^{(\alpha_i,\beta)}_{0,T_i,\vartheta}(t)$, $i = 1,\ldots,p$, and $\bar T_p = \sum_{i=1}^{p} T_i$. Then

$$\bar g_p(t) \to \frac{1}{\sigma_p}\mathrm{G}\left(\frac{t-\mu_p}{\sigma_p}\right) \quad \text{for} \quad p\to\infty$$

with μ_p and σ_p^2 the mean value and variance of the filter g_p. For their closed-form expressions the reader is referred to (Kiltz, 2017, Eq. (3.11)). It has also been proven there that for $p\to\infty$ the diff. $\bar g_p$ corresponds to the best compromise between the width of the impulse response and that of the frequency response, i.e., for $p\to\infty$, $\bar g_p$ converges to an optimal compromise in the sense of the uncertainty principle between the resolutions in the time domain and the frequency domain. This explains the observation made in Mboup and Riachy (2014): Two successive first-order diffs. instead of one second order diff. can be better.

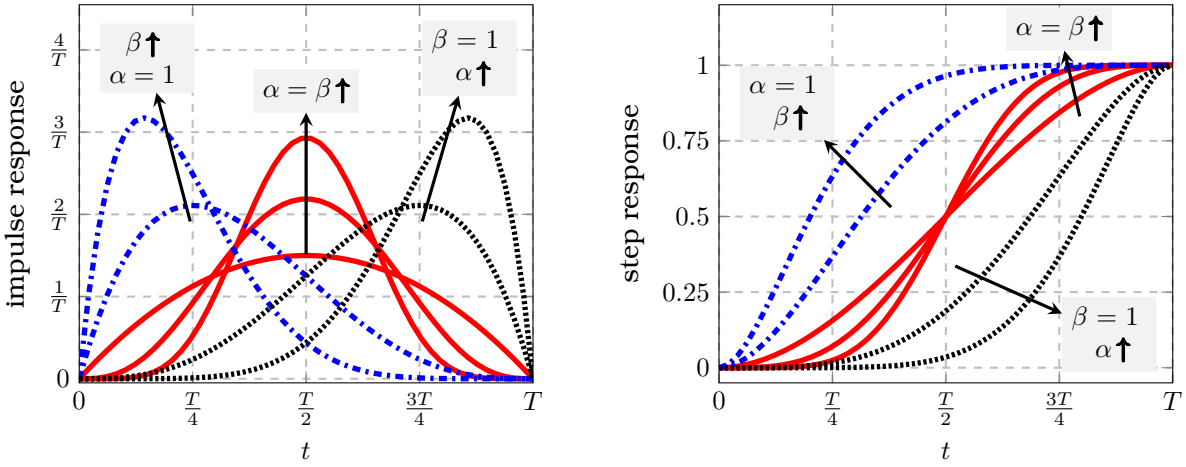


Figure 2.15: Impulse and step responses of a minimal Jac. diff. $g^{(\alpha,\beta)}_{N,T,\vartheta}$ with the parameters $\alpha, \beta \in \{1, 3, 6\}$, and $N = 0$.

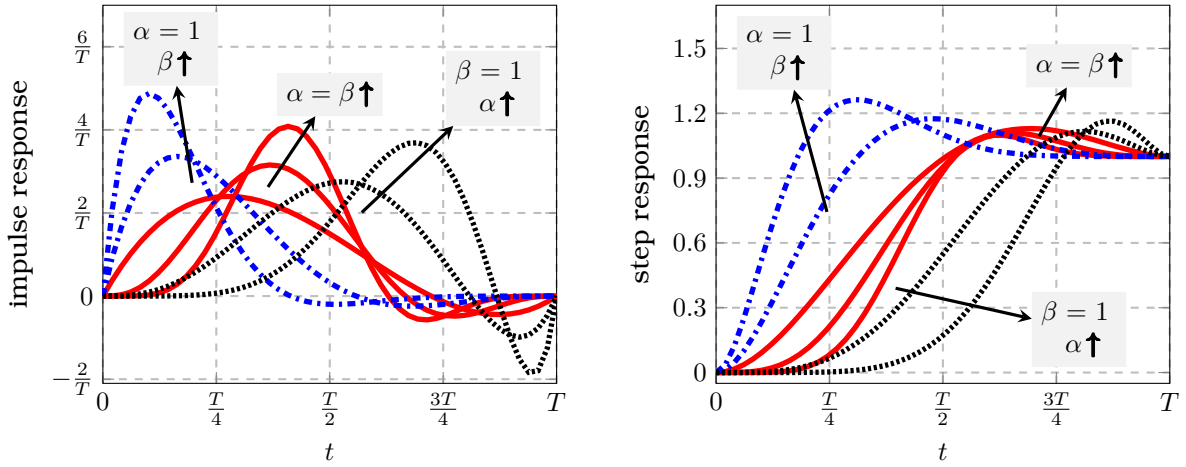


Figure 2.16: Impulse and step responses of a Jac. diff. $g^{(\alpha,\beta)}_{N,T,\vartheta}$ with the parameters $\alpha, \beta \in \{1, 3, 6\}$, $N = 1$, and $\vartheta = \mathrm{p}_2^{(\alpha,\beta)}$.

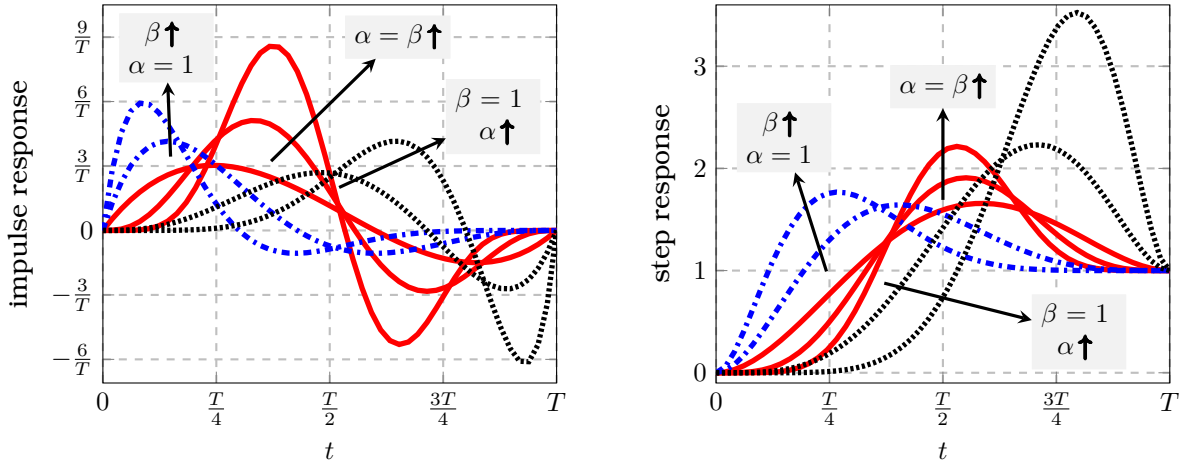


Figure 2.17: Impulse and step responses of a delay-free Jac. diff. $g^{(\alpha,\beta)}_{N,T,\vartheta}$ with the parameters $\alpha, \beta \in \{1, 3, 6\}$, $\vartheta = 1$, and $N = 1$.

Furthermore, it has been shown in (Kiltz, 2017, Sec. 3.2.2) that the diff. $g_{0,\bar{T}_p,\vartheta}^{(\bar{\alpha}_p,\bar{\beta}_p)}$ with

$$\bar{\alpha}_p = \left(\frac{\mu_p}{\sigma_p}\right)^2 \left(1 - \frac{\mu_p}{\bar{T}_p}\right) - \left(1 + \frac{\mu_p}{\bar{T}_p}\right) \quad \text{and} \quad \bar{\beta}_p = \left(\left(\frac{\mu_p}{\sigma_p}\right)^2 + 1\right)\left(\frac{\mu_p}{\bar{T}_p} - 2\right) + \frac{\bar{T}_p\mu_p}{\sigma_p^2}$$

is an approximation of a Gaussian kernel with mean value μ_p and variance σ_p^2.

Frequency-domain properties: Denote by $\omega \mapsto \mathcal{F}(\omega)$ the Fourier transform of the signal $t \mapsto f(t)$. Recall that the algebraic derivative approximation of the n-th order derivative of f is

$$\hat{f}^{(n)}(t) = \int_0^T g^{(n)}(\tau) f(t-\tau)\mathrm{d}\tau, \quad \text{with} \quad g^{(n)}(\tau) = \left(\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^n g_{N,T,\vartheta}^{(\alpha,\beta)}(\tau),$$

the Fourier transform of which reads

$$\mathcal{F}\left\{\hat{f}^{(n)}\right\}(\omega) = (\iota\omega)^n \mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\iota\omega)\mathcal{F}(\omega),$$

with $\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}$ the Fourier transform of $g_{N,T,\vartheta}^{(\alpha,\beta)}$. Again it is clear that this numerical differentiation approach falls within the framework of classical approaches for differentiation in the frequency domain, where a smoothing filter, in this case $\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}$, is followed by an ideal differentiation operator $(\iota\omega)^n$.

Analysing $\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}$ is thus crucial to tune the parameters. Its closed-form has been derived in Kiltz (2017); Kiltz, Mboup, and Rudolph (2012); Mboup and Riachy (2018). One possible representation of $\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}$ is

$$\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega) = \sum_{i=0}^{N} \frac{(\alpha+\beta+2i+1)\mathrm{P}_i^{(\alpha,\beta)}(\vartheta)}{\alpha+\beta+i+1} \sum_{k=0}^{i} (-1)^{i-k} \binom{i}{k} \mathrm{M}_{i,k}^{(\alpha,\beta)}(-\iota\omega T)$$

with

$$\mathrm{M}_{i,k}^{(\alpha,\beta)}(z) = \mathrm{M}(\alpha+i-k+1, \alpha+\beta+i+2, z), \quad z \in \mathbb{C},$$

where M is the confluent hypergeometric function defined in Abramowitz and Stegun (1965). Since $\omega \mapsto \mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)$ is continuously differentiable it follows that the Fourier transform $\mathcal{H}$ of h with h defined in (2.38) also has this property. It is then easy to show that the admissibility condition in (2.39) is satisfied. Thus, the wavelet interpretation given Section 2.3.5 is valid for Jacobi polynomials, which has been first discovered in Mboup and Riachy (2018).

In the frequency domain, the symmetry property of $g_{N,T,\vartheta}^{(\alpha,\beta)}$ discussed earlier and given in (2.69) corresponds to

$$\mathrm{e}^{\iota\frac{\omega T}{2}} \mathcal{G}_{N,T,\vartheta}^{(\beta,\alpha)}(\omega) = \mathrm{e}^{-\iota\frac{\omega T}{2}} \mathcal{G}_{N,T,-\vartheta}^{(\alpha,\beta)}(-\omega)$$

and implies

$$\left|\mathcal{G}_{N,T,\vartheta}^{(\beta,\alpha)}(\omega)\right| = \left|\mathcal{G}_{N,T,-\vartheta}^{(\alpha,\beta)}(\omega)\right|.$$

As in Kiltz (2017); Mboup and Riachy (2018), it is assumed in the sequel that $\alpha \leqslant \beta$. By the latter symmetry property the next results hold also for $\alpha > \beta$ by exchanging α and β and mirroring the estimation delay at $T/2$.

Annihilators

For the estimation of the n-th order derivative, $n \in \{0, \ldots, n^\star\}$, the operational domain formulation in (2.32) is reconsidered. All the terms $s^{N-i}x^{(i)}(0)$, $i \neq n$, are now viewed as undesired

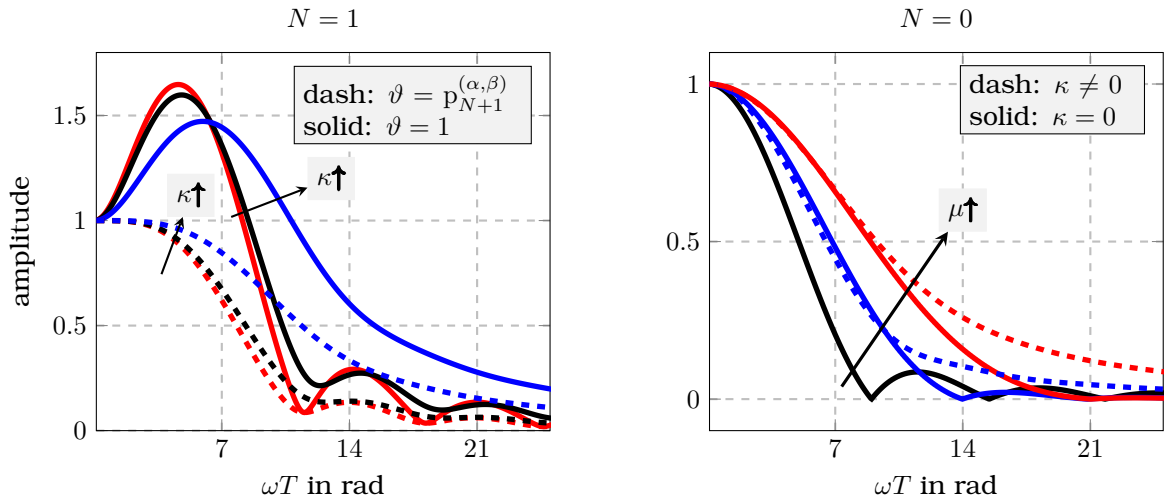


Figure 2.18: Variation of the amplitude spectrum $\omega \mapsto \left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ of a diff. $g_{N,T,\vartheta}^{(\alpha,\beta)}$ for fixed N and various values of $\kappa = |\alpha - \beta|$, $\mu = 1 + \min\{\alpha, \beta\}$, and ϑ.

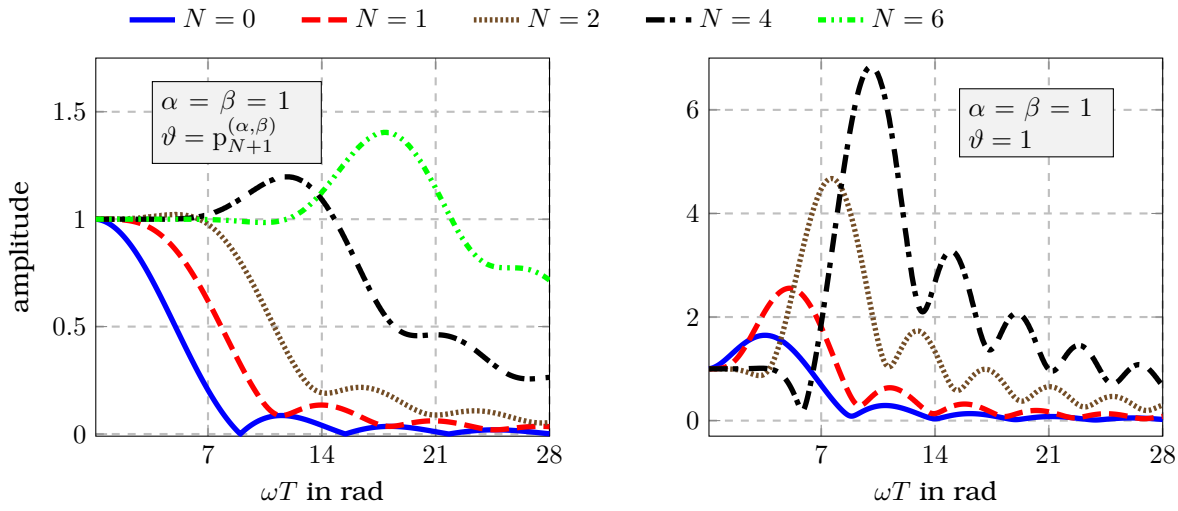


Figure 2.19: Variation of the amplitude spectrum $\omega \mapsto \left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ of a diff. $g_{N,T,\vartheta}^{(\alpha,\beta)}$ for fixed parameters α and β and various values of N.

perturbations to be annihilated. Assume first that the Taylor series (2.32) is truncated at the sought derivative, i.e. $n^\star = n$.

Consider now the differential operator

$$\Pi_n^{(\alpha_0,\beta_0)} = \frac{1}{s^{N+\beta_0+1}} \left(\frac{\mathrm{d}}{\mathrm{d}s}\right)^{n+\alpha_0} s^n, \quad \alpha_0, \beta_0 \in \mathbb{N},$$

associated to the rational function

$$\rho_n^{(\alpha_0,\beta_0)}(s) = \frac{(-1)^{n+\alpha_0}(n+\alpha_0)!}{s^{\alpha_0+\beta_0+2+2n}}.$$

The notation $\hat{f}^{(n)}(0; \alpha_0, \beta_0, T)$ is adopted for the approximation of $f^{(n)}(0)$ when $n^\star = n$. Transforming

$$\Pi_n^{(\alpha_0,\beta_0)} \mathcal{X}_N(s) = \rho_n^{(\alpha_0,\beta_0)}(s) f^{(n)}(0),$$

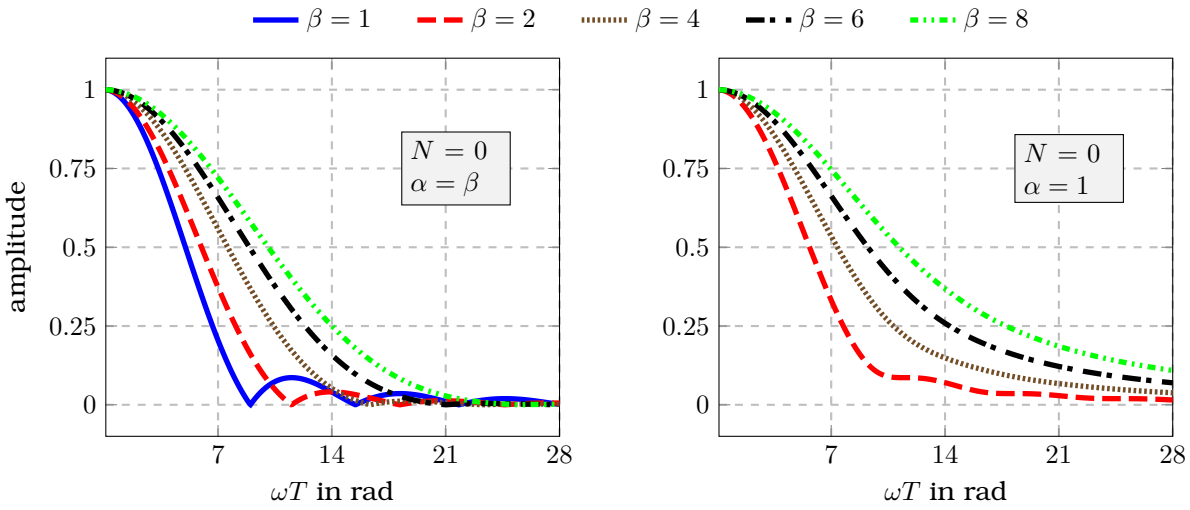


Figure 2.20: Variation of the amplitude spectrum $\omega \mapsto \left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ of a diff. $g_{N,T,\vartheta}^{(\alpha,\beta)}$ for $N = 0$ and various values of α and β.

back into the time domain yields the expression

$$\hat{f}^{(n)}(0;\alpha_0,\beta_0,T) = \frac{\gamma_{\alpha_0,\beta_0,n}}{T^{\alpha_0+\beta_0+2n+1}} \int_0^T (T-\tau)^{\beta_0+n} \tau^{\alpha_0+n} f^{(n)}(\tau) \mathrm{d}\tau,$$

where T denotes an estimation time and

$$\gamma_{\alpha_0,\beta_0,n} = \frac{(\alpha_0+\beta_0+2n+1)!}{(\alpha_0+n)!(\beta_0+n)!}.$$

To obtain a causal approximation of the n-th order derivative at a time $t \geqslant T$, it is sufficient to consider the sliding window $\mathcal{I}_T(t) = [t-T, t]$, with window length T. This yields linear time-invariant filters and the resulting approximated derivative corresponds to

$$\hat{f}^{(n)}(t) = \int_0^T g_{0,T,\vartheta}^{(\alpha_0+n,\beta_0+n)}(\tau) f^{(n)}(t-\tau) \mathrm{d}\tau.$$

with $g_{0,T,\vartheta}^{(\alpha_0+n,\beta_0+n)}$ from (2.65b). Thus, the Jac. diff. with $\alpha, \beta \in \mathbb{N}$ and $N = 0$ can be derived using algebraic manipulations of truncated Taylor series.

Recall from (2.68) that any Jac. diff. with $N > 0$ can be written as a linear combination of minimal Jac. diffs., i.e., diff. with $N = 0$. As for the case of Laguerre polynomials in Section 2.4.2, a differential operator can be derived to get the approximation of the derivative. Since the steps are identical the details are omitted here. The original derivation of the operators for $N > 0$ has been proposed in Mboup (2009); Mboup et al. (2007).

2.4.4 Bibliographical and historical comments

In[18] Fliess and Sira-Ramírez (2003a) new algebraic approaches for the non-asymptotic estimation of parameters and states of linear systems have been proposed. These approaches are based on module theory, differential algebra, and operational calculus. They permit the

[18]These bibliographical and historical comments related to the Jac. diffs. stem entirely from Othmane et al. (2022). The reader is referred to this reference for a more detailed discussion and an exhaustive survey of applications of Jac. diff.

annihilation of structured perturbations and exhibit good robustness properties with respect to corrupting disturbances, without the need to know their statistical properties, as discussed in Fliess (2006, 2008).

The approaches have then been used and extended in Fliess, Join, Mboup, and Sira-Ramírez (2004); Fliess, Join, and Mounier (2004); Fliess and Sira-Ramírez (2004) for the estimation of time derivatives of noisy signals. The works Mboup (2009); Mboup et al. (2007); Mboup et al. (2009) have further extended these approaches to obtain a least squares interpretation for the approximated derivatives and an implementation in terms of linear shift-invariant finite impulse response (FIR) digital filters. With these filters the estimation of individual derivatives is possible, which avoids the inversion of a possibly ill-conditioned matrix, as discussed in Mboup et al. (2007). Each of these individual derivatives can be seen as a single parameter to be estimated. In Mboup et al. (2009) it has been shown that admitting a small delay in the approximation can improve its accuracy by one order, where this favourable delay can be computed from the parameters of the diffs. Later works Kiltz (2017); Kiltz and Rudolph (2013); Liu et al. (2010, 2011a, 2011b) have derived the filters in the time domain by using orthogonal projections and Jacobi polynomials. This has extended the range of the parameters involved from the set of natural numbers to a subset of the real numbers.

The use of the Gaussian kernel from (2.44) as a mollifier has been discussed in detail in (Murio, 2011, Ch. 1) and its use in the control community has been described in Braci and Diop (2003); Diop, Fromion, and Grizzle (2001), for example. However, the resulting estimation delay has not been discussed in earlier works. The use of Hermite polynomials allows richer signals models and decreases the estimation delay. An algebraic derivation of annihilators for Hermite diffs. and Lag. diffs. using Weyl algebra has been proposed in Ushirobira (2018); Ushirobira and Quadrat (2016). The properties of the resulting diffs., like the estimation error, have been discussed for the first time in this chapter.

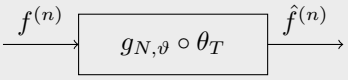
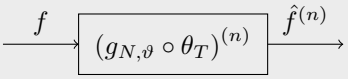
2.5 Summary and concluding remarks

Different derivation methods and interpretations for numerical differentiation approaches based on OPs have been discussed in this chapter. The differential-algebraic manipulations of truncated Taylor series in the algebraic context has shown the relation to the linear identifiability of parameters. The derivation using the orthogonal projections has been used for the analysis of the estimation errors and the estimation delay. The filter interpretation has permitted to associate to these diffs. a system theoretic interpretation. Table 2.4 summarises these findings[19]. For all discussed approaches it has been shown that admitting a small but known delay reduces the estimation error by one order and avoids overshoots in the amplitude spectra.

All orth. diffs. discussed in this chapter depend on several parameters. During their design for specific applications, it is important to choose the parameters such that the estimation error is minimised. Needless to say that finding good compromises without relying on trial-and-error approaches is important. Moreover, in most applications, measurements are available at discrete sampling instants only. Thus, the convolution integral (2.35) must be approximated

[19] A similar table summarizing the properties of Jac. diffs. has first been published in Othmane et al. (2022)

Table 2.4: Summary of the interpretations of the discussed diffs. $g_{N,\vartheta}$ from Definition 2.5 and the practical usage of each.

Context	Interpretation	Practical usage
Algebraic	Annihilation of undesired perturbations with the annihilator in (2.33)	Relation to linear identifiability
Approximation theoretic	Polynomial approximation of derivative using a Hilbert reproducing kernel	Analysis of estimation properties and relation to established differentiation methods
Systems theoretic	Linear time invariant filtering of the sought derivative in (2.36): 	Analysis of filter properties and parametrisation
	Linear time invariant filtering of the measured signal in (2.35): 	Implementation

by an appropriate quadrature method. These problems are addressed in detail in the next chapter.

3

Tuning and real-time implementation of classical orthogonal differentiators

The differentiators (diffs.) developed in the last chapter depend on numerous parameters. For example, a Jacobi differentiator (Jac. diff.) depends on the parameters α, β, T, ϑ, and N. Thus, five numerical values must be assigned to these parameters, which is challenging since the approximated derivative cannot be compared to the true one. For instance, consider the parameter T: Decreasing it decreases the estimation error stemming from the polynomial approximation and increases the noise contribution, whereas increasing T yields the opposite results (see, e.g., Corollaries 2.4 and 2.5). Finding good compromises without relying on trial-and-error approaches is important in practice. Furthermore, in most applications, the measurement is only available at discrete sampling instants. Thus, the convolution integrals seen in the last chapter must be approximated by sums.

This chapter discusses systematic tuning guidelines for the diffs. based on classical orthogonal polynomials. The guidelines can be used to design diffs. with desired frequency-domain properties. The discrete-time implementation is then addressed in light of these properties. Different discretisation schemes are discussed, and the preservation of the frequency-domain properties is considered using an error norm initially proposed for Jac. diff. in Kiltz (2017). It is shown that all diffs. can be implemented as finite-impulse response (FIR) filters, with a filter window length that needs to be computed. It is then shown that Jacobi and Laguerre diffs. can be efficiently implemented using recursive schemes. For the Jac. diffs., the recursive implementation has been derived in Kiltz (2017). This implementation can significantly reduce the memory requirements and computational burden. Moreover, it is shown that Laguerre differentiators (Lag. diffs.) can be efficiently used for the simultaneous approximation of numerous derivatives. This bridges the gap to known approaches like the high-gain (HG) diffs. (see, e.g., Dabroom and Khalil (1997, 1999); Esfandiari and Khalil (1992)) and the state variable filters (see, e.g., Hofmann, Lion, and Best (1966); Peter and Isermann (1990); Young (1981)). It is shown that for a special parametrisation of the Lag. diffs. it is possible to recover state variable filters and HG diffs.

The parametrisation of Jac. diffs. has first been discussed in Mboup et al. (2007) and then in Liu (2011); Liu et al. (2010, 2011a) in the time-domain. Later works Kiltz (2017); Kiltz and Rudolph (2013); Mboup and Riachy (2014, 2018) propose frequency-domain tuning guidelines. A summary of these discussions is provided in Othmane et al. (2022). Thus, this chapter only summarises the most important tuning guidelines for the Jac. diffs. While

Laguerre and Hermite polynomial-based diffs. have been proposed in Ushirobira (2018); Ushirobira and Quadrat (2016), no tuning guidelines have been developed and the discrete-time implementation has not been addressed in the literature.

The systematic tuning guidelines and the discretisation issues are illustrated using some academic examples first. Then, different experimental case studies are performed to show the application of the diffs. The first and second order derivatives of a known but disturbed laboratory measurement are approximated. The three diffs. are compared with respect to the quality of the estimation and the required computational burden and storage requirements. Then, the use of diffs. for the approximate inversion of analogue anti-aliasing filters is discussed.

The chapter ends with Section 3.4 discussing an automatic tuning approach for all diffs. This approach is based on an optimisation problem that requires only the measured signal. It is shown that the approach yields excellent results with experimental data.

Chapter content

3.1 Systematic tuning of classical orthogonal differentiators 54
3.1.1 Tuning of Hermite differentiators 54
3.1.2 Tuning of Laguerre differentiators 55
3.1.3 Tuning of Jacobi differentiators 56
3.2 Implementation of classical orthogonal differentiators 59
3.2.1 Implementation of Hermite differentiators 60
3.2.2 Implementation of Laguerre differentiators 64
3.2.3 Implementation of Jacobi differentiators 72
3.3 Application of systematic tuning guidelines 73
3.3.1 Derivative approximation 74
3.3.2 Approximate inversion of analogue anti-aliasing filters 82
3.4 Automatic tuning of orthogonal differentiators 84
3.4.1 Proposed approach 85
3.4.2 Examples for automatic tuning 87
3.4.3 Concluding remarks 90
3.5 Summary and concluding remarks 92

3.1 Systematic tuning of classical orthogonal differentiators

In the following, systematic tuning guidelines for the parametrisation of the diffs. shall be discussed. These approaches rely on the analysis of the Fourier transform of the diffs.

3.1.1 Tuning of Hermite differentiators

From the analysis of the Fourier transform of Hermite differentiators (Hermite diffs.) in Section 2.4.1 it can be seen that in the amplitude of the Fourier transform given in (2.46) the weight function of the Hermite polynomials appears. It is then easy to conclude that the Hermite diffs. have a low-pass property: While the low-frequency components of signals are

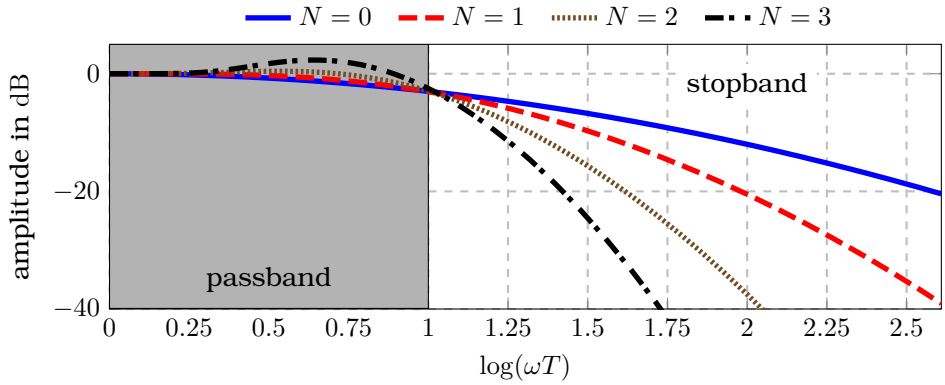


Figure 3.1: Low-pass interpretation of Hermite diffs. with delay for different values of the parameter N. The cutoff frequency is the frequency for which the amplitude is equal to $-3\,\mathrm{dB}$.

passed through, high-frequency ones are attenuated. In the case of a delay-free approximation, high overshoots can, however, appear in the amplitude spectrum, as shown in Fig. 2.4.

Thus, for a delayed approximation, the parameter T can be adapted to achieve, for frequencies ω that are lower than a specific frequency ω_c, an amplitude attenuation lower than F_{ω_c}. The frequency ω_c and the interval $|\omega| \leqslant \omega_\mathrm{c}$ are called the cutoff frequency and the passband, respectively. The interval $|\omega| > \omega_\mathrm{c}$ is called stopband. This low-pass filter interpretation is illustrated in Fig. 3.1. The parameter μ does not affect the amplitude spectrum and only influences the phase spectrum as can be seen in (2.46), for example.

Hermite diffs. with $N = 0$, also known as Gaussian filters (see, e.g., Papoulis (1962)), have been very popular to filter disturbed measurements. Application of these low-pass filters can be found in Canny (1986); Koenderink (1984); Marr and Hildreth (1980); Witkin (1987) for image processing or Diop et al. (2001, 2000); Murio (2011) for numerical differentiation to name only a few.

3.1.2 Tuning of Laguerre differentiators

Recalling the amplitude spectrum of Lag. diffs. depicted in Figs. 2.11 and 2.12 for some example parametrisations, it can be observed that for low frequencies the amplitude is close to 0 dB and decreases significantly for high frequencies. Thus, a piecewise approximation of the amplitude spectrum of (2.59) that yields a low-pass interpretation shall be proposed.

The limit

$$\lim_{\omega \to 0} \mathcal{F}\left\{\bar{g}^{(\alpha)}_{N,T,\vartheta}\right\}(\omega) = 1$$

and the expansion[1]

$$\mathcal{F}\left\{\bar{g}^{(\alpha)}_{N,T,\vartheta}\right\}(\omega) = \frac{1}{(\imath\omega T)^{\alpha+1}} \sum_{n=0}^{N} \mathrm{L}_n^{(\alpha)}(\vartheta) + \mathcal{O}\left(\frac{1}{(\imath\omega T)^{\alpha+2}}\right), \quad \omega \to \infty, \tag{3.1a}$$

[1]The expansion (3.1a) follows from the fact that

$$\frac{1}{(\imath\omega T + 1)^{\alpha+1}} = \frac{1}{(\imath\omega T)^{\alpha+1}} \left(1 + \frac{1}{\imath\omega T}\right)^{-\alpha-1} = \frac{1}{(\imath\omega T)^{\alpha+1}} \sum_{n=0}^{\infty} \binom{-\alpha-1}{n} \left(\frac{1}{\imath\omega T}\right)^n,$$

with $\binom{-\alpha-1}{n} = (-\alpha-1)(-\alpha-2)\ldots(-\alpha-n)/n!$. The binomial series converges for $|\omega| > T$ (see, e.g., Abel (1826)).

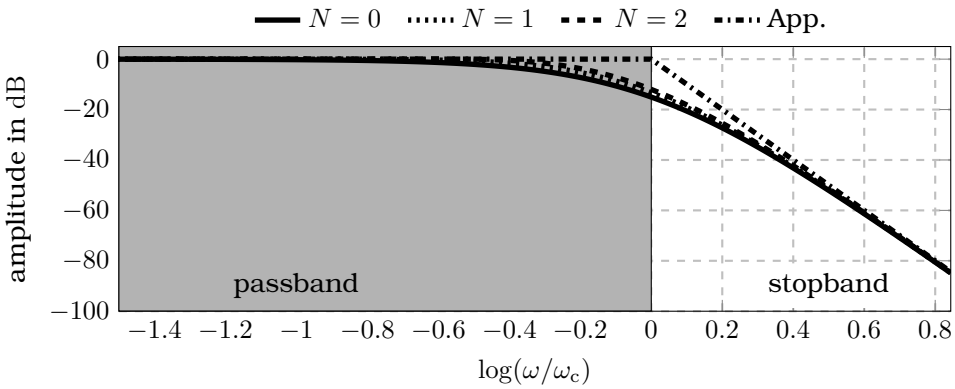


Figure 3.2: Amplitude spectra of Lag. diffs. with delay for various values of N, $\alpha = 4$ and $\vartheta = l_{N+1}^{(\alpha)}$. The approximation (3.2) is depicted with a dashed line.

motivate the piecewise approximation

$$\left|\mathcal{F}\left\{\bar{g}_{N,T,\vartheta}^{(\alpha)}\right\}(\omega)\right| \approx \begin{cases} 1, & \text{for} \quad |\omega| < \omega_c, \\ |\frac{\omega}{\omega_c}|^{-\alpha-1}, & \text{otherwise}, \end{cases} \tag{3.2a}$$

of the amplitude spectrum with

$$\omega_c = \frac{1}{T}\left|\sum_{n=0}^{N} \mathrm{L}_n^{(\alpha)}(\vartheta)\right|^{\frac{1}{\alpha+1}}. \tag{3.2b}$$

The frequency ω_c is called the cuttoff frequency of the diff. The intervals $|\omega| \leqslant \omega_c$ and $|\omega| > \omega_c$ are called the passband and the stopband, respectively. The amplitude spectrum asymptotically decreases by $20(\alpha+1)\,\mathrm{dB}$ per frequency decade for high frequencies. Thus, the diff. can be interpreted as a low-pass filter of order $\alpha+1$. Figs. 3.2 to 3.4 depict the amplitude spectra of diffs. for various parameter combinations and the corresponding approximations from (3.2). It can be seen that it well describes the relevant properties of the amplitude spectra except around overshoots that appear for the delay-free parametrisation as shown in Fig. 3.4.

This analysis motivates the following parametrisation procedure. First, choose the polynomial degree N and the order of the low-pass filter $\alpha+1$, i.e., the stopband slope. Then, the parameter T can be calculated to achieve a desired cutoff frequency ω_c by rewriting (3.2b) as

$$T = \frac{1}{\omega_c}\left|\sum_{n=0}^{N} \mathrm{L}_n^{(\alpha)}(\vartheta)\right|^{\frac{1}{\alpha+1}}.$$

3.1.3 Tuning of Jacobi differentiators

First tuning aspects of Jac. diffs. have been discussed in the initial works Mboup et al. (2007); Mboup et al. (2009). Further discussions on the effects of the parameters in the time domain can be found in Liu (2011); Liu et al. (2010); Liu et al. (2008). Promising tuning approaches based on the frequency-domain analysis of the diffs. have been proposed in Kiltz and Rudolph (2013); Mboup and Riachy (2014, 2018); Othmane, Mounier, and Rudolph (2021). The following paragraphs have first been published in Othmane et al. (2022) and recall the most important approaches that are useful for a systematic tuning of the diffs. to achieve desired frequency-domain properties.

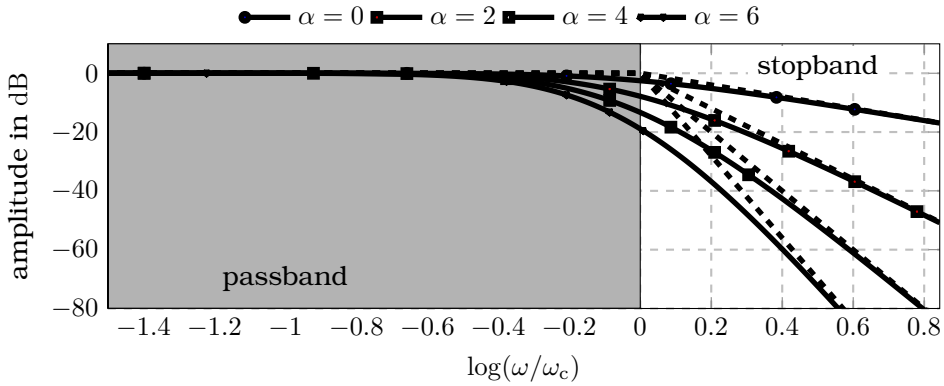


Figure 3.3: Amplitude spectra of Lag. diffs. with delay for varying filter orders α, $N = 1$ and $\vartheta = l_{N+1}^{(\alpha)}$. Dashed lines show the approximations from (3.2).

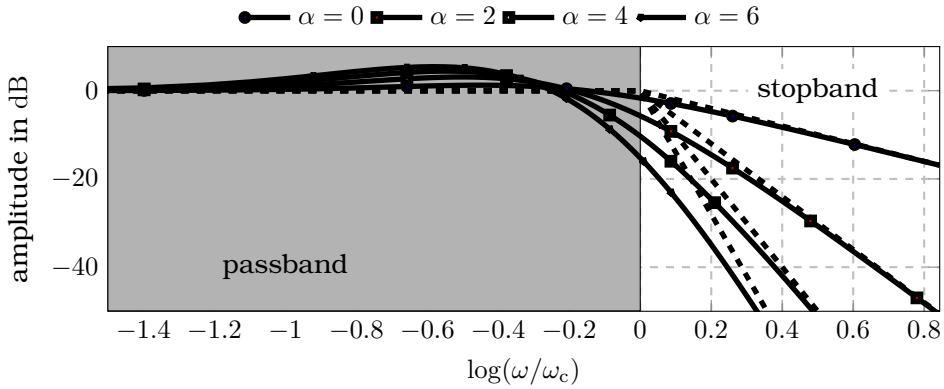


Figure 3.4: Amplitude spectra of delay-free Lag. diffs. for varying filter orders α and parameters $N = 1$ and $\vartheta = 0$. The approximations from (3.2) are given in dashed lines.

Exact analysis

The effect of N on the filter amplitude spectrum is considered first. While increasing N augments the accuracy of the underlying signal model, it increases the sensitivity to noise. It has been shown in (Mboup & Riachy, 2018, Prop. 3.1) that the $\mathcal{L}_2$-norm $\left\|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}\right\|_{\mathcal{L}_2}$ is an increasing function of N. Thus, the filter output noise power increases with N. The variation of $\left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ with respect to N is depicted in Fig. 2.19. For the delay-free parametrisation significant overshoots can be observed. This explains the poor performance of the delay-free Jac. diff. noticed in Mboup et al. (2009) and explained in Mboup and Riachy (2014).

Minimal diffs. allow for a particularly pronounced noise rejection. It has been proven in (Mboup & Riachy, 2018, Lem. 3.1) that

$$\left\|\mathcal{G}_{0,T,\vartheta}^{(\alpha+l,\beta)}\right\|_{\mathcal{L}_2} < \left\|\mathcal{G}_{0,T,\vartheta}^{(\alpha+m,\beta)}\right\|_{\mathcal{L}_2}, \quad l < m,$$

i.e., the overall noise power increases with the difference $|\alpha - \beta|$ as depicted in Fig. 2.20. This effect has first been observed in Mboup and Riachy (2014) and has led the authors in Mboup and Riachy (2018) to recommend the parametrisation $\alpha = \beta$ also for the affine (and minimum variance) diffs. .

For $\alpha = \beta$ and $N = 0$ the amplitude spectrum possesses infinitely many zeros. In fact, as

discussed in Kiltz (2017); Kiltz et al. (2012) the Fourier transform reads

$$\mathcal{G}_{0,T,\vartheta}^{(\alpha,\alpha)}(\omega) = \mathrm{e}^{-\iota\frac{\omega T}{2}}\Gamma\left(\alpha+\frac{3}{2}\right)\left(\frac{4}{\omega T}\right)^{\alpha+\frac{1}{2}}\mathrm{J}_{\alpha+\frac{1}{2}}\left(\frac{\omega T}{2}\right),$$

where $\mathrm{J}_{\alpha+\frac{1}{2}}$ is the Bessel function of the first kind and order $\alpha+1/2$ defined in #A.20. The parameters α and T can now be chosen such that the zeros of $\mathcal{G}_{0,T,\vartheta}^{(\alpha,\alpha)}$ coincide with specific frequencies. This has been used in Kiltz (2017); Kiltz, Janocha, and Rudolph (2013); Kiltz and Rudolph (2013) to compute derivatives of experimental signals corrupted by a harmonic disturbance with a known frequency without resorting to further filtering methods. The approach is extended in Othmane, Mounier, and Rudolph (2021) to approximately annihilate disturbances with Fourier transforms showing repetitive peaks and applied to the estimation of derivatives of quantised signals.

Asymptotic analysis

The work Kiltz and Rudolph (2013) shows that algebraic diffs. can be interpreted as low-pass filters with a desired cutoff frequency and arbitrary low-pass order, i.e., stopband slope. This will be briefly reviewed along with the extensions in Kiltz (2017).

Using the asymptotic expansion of the confluent hypergeometric function (Abramowitz & Stegun, 1965, Ch. 13), it has been proven in (Kiltz, 2017, Sec. B.3.3) that the amplitude spectrum $\omega \mapsto \left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ of the diff. $g_{N,T,\vartheta}^{(\alpha,\beta)}$ satisfies

$$\left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right| = \left|\widehat{\mathcal{G}}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right| + \mathcal{O}\left(|\omega|^{-\mu-\min\{\kappa,1\}}\right), \quad \omega\to\infty,$$

where

$$\left|\widehat{\mathcal{G}}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right| = \frac{q_{N,\vartheta}^{(\alpha,\beta,\sigma)}}{\Gamma(\mu+\kappa)\,|\omega T|^{\mu}}, \quad \mu = 1+\min\{\alpha,\beta\}, \quad \kappa = |\alpha-\beta|, \tag{3.3}$$

and

$$q_{N,\vartheta}^{(\alpha,\beta,\sigma)} = \begin{cases} \Gamma(\mu)\max\left\{\left|r_{N,T}^{(\mu,0,\sigma)}\right|, s_{N,T}^{(\mu,0,\sigma)}\right\}, & \kappa=0,\\ \Gamma(\mu+\kappa)\left|r_{N,T}^{(\mu,\kappa,\sigma)}\right|, & \kappa>0,\end{cases}$$

$$\sigma = \begin{cases} 1, & \alpha\leqslant\beta,\\ -1, & \alpha>\beta,\end{cases}$$

$$r_{N,\vartheta}^{(\mu,\kappa,\sigma)} = \sum_{i=0}^{N}\frac{c_i^{(\mu,\kappa)}}{\Gamma(\mu+\kappa+i)}\mathrm{P}_i^{(\mu-1,\mu+\kappa-1)}(\sigma\vartheta),$$

$$s_{N,\vartheta}^{(\mu,\kappa,\sigma)} = \sum_{i=0}^{N}(-1)^i\frac{c_i^{(\mu,\kappa)}}{\Gamma(\mu+i)}\mathrm{P}_i^{(\mu-1,\mu+\kappa-1)}(\sigma\vartheta),$$

$$c_i^{(\mu,\kappa)} = (2\mu+\kappa+2i-1)\,\Gamma(2\mu+\kappa+i-1).$$

Consequently, for large frequencies the amplitude spectrum $\omega\mapsto\left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ can be approximated by the asymptote $\omega\mapsto\left|\widehat{\mathcal{G}}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$.

Additionally, in (Kiltz, 2017, Sec. B.3.3) the asymptotes

$$\underline{\widehat{\mathcal{G}}_{N,T,\vartheta}^{(\alpha,\beta)}}(\omega) = \left|\left|\frac{r_{N,\vartheta}^{(\mu,\kappa,\sigma)}}{(\omega T)^{\mu}}\right| - \left|\frac{s_{N,\vartheta}^{(\mu,\kappa,\sigma)}}{(\omega T)^{\mu+\kappa}}\right|\right|, \quad \overline{\widehat{\mathcal{G}}_{N,T,\vartheta}^{(\alpha,\beta)}}(\omega) = \left|\frac{r_{N,\vartheta}^{(\mu,\kappa,\sigma)}}{(\omega T)^{\mu}}\right| + \left|\frac{s_{N,\vartheta}^{(\mu,\kappa,\sigma)}}{(\omega T)^{\mu+\kappa}}\right| \tag{3.4}$$

for the lower and upper envelopes, respectively, of the amplitude spectrum of the diff. have been derived.

For $N = 0$ and $\alpha = \beta$, it holds that $\left|\underline{\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}}(\omega)\right| = 0$, which is consistent with the observation that the filter amplitude spectrum has infinitely many zeros.

Fig. 3.5 depicts two amplitude spectra with the corresponding approximations. It is clear that even though the latter approximation is derived for large frequencies, the convergence rate of the approximation error is high and the bounds can be used for the in depth of the effects of the parameters on the filter. In Othmane, Rudolph, and Mounier (2021a) a region of validity for the approximation is derived.

The function $\omega \mapsto \Delta\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega) = \overline{\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}}(\omega) - \underline{\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}}(\omega)$ describes the distance between the bounds introduced in (3.4). It satisfies

$$\Delta\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega) = p^{(\alpha,\beta,\sigma)}_{N,T} \left|\frac{\omega_c}{\omega}\right|^{\mu+\kappa}, \tag{3.5}$$

for $\kappa = 0$ and $|\omega T|^\kappa > k = \left|r^{(\mu,\kappa,\sigma)}_{N,\vartheta}/s^{(\mu,\kappa,\sigma)}_{N,\vartheta}\right|$. In the latter equation

$$\begin{aligned} p^{(\alpha,\beta,\sigma)}_{N,T} &= \begin{cases} 2\min\{k, 1/k\}, & \text{for} \quad \kappa = 0, \\ 2k/\left|r^{(\mu,\kappa,\sigma)}_{N,\vartheta}\right|^\kappa, & \text{otherwise}, \end{cases} \\ \omega_c &= \frac{1}{T}\left(\frac{q^{(\alpha,\beta,\sigma)}_{N,\vartheta}}{\Gamma(\mu+\kappa)}\right)^{\frac{1}{\mu}}, \end{aligned} \tag{3.6}$$

where ω_c is referred to as the cutoff frequency.

Comparing (3.6) with the asymptote $\left|\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega)\right|$ introduced in (3.3), it can be seen that

$$\left|\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega)\right| = \left|\frac{\omega_c}{\omega}\right|^\mu.$$

Since, moreover, it can be shown that $\lim_{\omega\to 0} \mathcal{G}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega) = 1$, Kiltz and Rudolph (2013) has proposed the following approximation of the amplitude spectrum of the Jac. diff.

$$\left|\mathcal{G}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega)\right| \approx \widetilde{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega) = \begin{cases} 1, & \text{for} \quad |\omega| \leqslant \omega_c, \\ \left|\frac{\omega_c}{\omega}\right|^\mu, & \text{otherwise}. \end{cases} \tag{3.7}$$

As it can be seen in Fig. 3.5 for two example diffs., (3.7) is a remarkably good approximation of the amplitude spectrum, except around overshoots.

The tuning guidelines developed in Kiltz and Rudolph (2013) can then be easily deduced from (3.7): The diff. window length T can be chosen by using a desired cutoff frequency from (3.6). The parameter $\mu = 1 + \min\{\alpha, \beta\}$ can be used to get a desired filter order, i.e., the stopband slope given by $20\mu\,\mathrm{dB}$. It follows from (3.5) that increasing $\kappa = |\alpha - \beta|$ increases the convergence speed of the difference $\Delta\widehat{\mathcal{G}}^{(\alpha,\beta)}_{N,T,\vartheta}(\omega)$ between the bounds to zero, which supports the statement from Mboup and Riachy (2018) reported above that the choice $\alpha = \beta$ decreases the sensitivity to noise.

3.2 Implementation of classical orthogonal differentiators

In most applications, the function f is known at discrete sampling instants only. Then, the convolution integrals must be approximated by an appropriate quadrature method, which

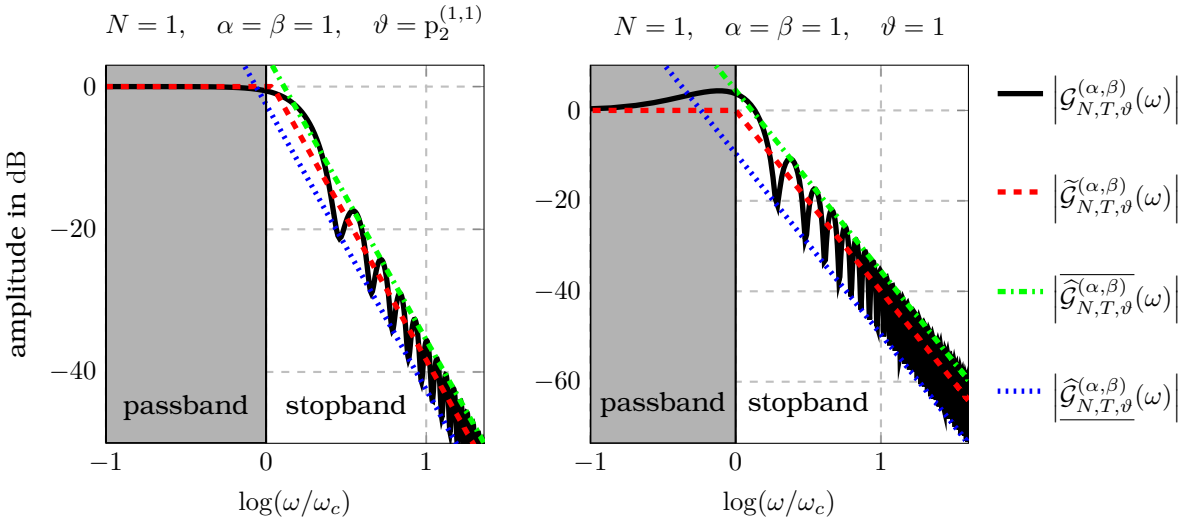


Figure 3.5: Amplitude spectrum $\omega \mapsto \left|\mathcal{G}_{N,T,\vartheta}^{(\alpha,\beta)}(\omega)\right|$ of a Jac. diff. $g_{N,T,\vartheta}^{(\alpha,\beta)}$ along with the upper and lower bounds from (3.4), the approximations $\widetilde{\mathcal{G}}_{N,T,\vartheta}^{(\alpha,\beta)}$ from (3.7) and the low-pass interpretation for two parametrisations. The left figure shows a diff. with delay. The right figure shows a delay-free one.

yields discrete FIR filters. In the following, equidistant sampling with sampling period t_s is assumed for simplicity. For the sake of brevity, the abbreviation $f_i = f(it_\mathrm{s})$, $i \in \mathbb{N}$, for a sample of a function f at time it_s is used in the sequel. The discretisation of the Hermite, Laguerre, and Jacobi diffs. is discussed in the following.

3.2.1 Implementation of Hermite differentiators

Recall from Section 2.4.1 that the n-th order derivative of a function f can be approximated as

$$\hat{f}^{(n)}(t) = \int_{-\infty}^{t} g^{(n)}(t-\tau) f(\tau) \mathrm{d}\tau, \quad g = g_{N,T,\vartheta,\mu}, \tag{3.8}$$

with $g_{N,T,\vartheta,\mu}$ a Hermite diff. defined in Definition 2.6.

Continuous-time FIR approximation

The values $g_{N,T,\vartheta,\mu}(\tau)$ vanish for $\tau \gg \mu$ and $\tau \ll \mu$ as depicted in Fig. 2.2 for different diffs., for example. This motivates the following approximation

$$\hat{f}^{(n)}(t) \approx \int_{\mu-\tau^-}^{\mu+\tau^+} g^{(n)}(\tau) f(t-\tau) \mathrm{d}\tau, \quad \tau^-, \tau^+ > 0,$$

where the integration bounds from (3.8) have been adapted. Thus, an FIR filter is recovered with the window length $T_L = \tau^+ + \tau^-$.

To determine τ^- and τ^+ the approach proposed in Kiltz (2017) for the truncation of the filter window length of Jac. diff. is used. Consider the distribution function

$$\Phi(t) = \frac{\int_{-\infty}^{t} \left|g^{(n)}(\tau)\right| \mathrm{d}\tau}{\mathrm{m}_0\left(\left|g^{(n)}\right|\right)}, \tag{3.9}$$

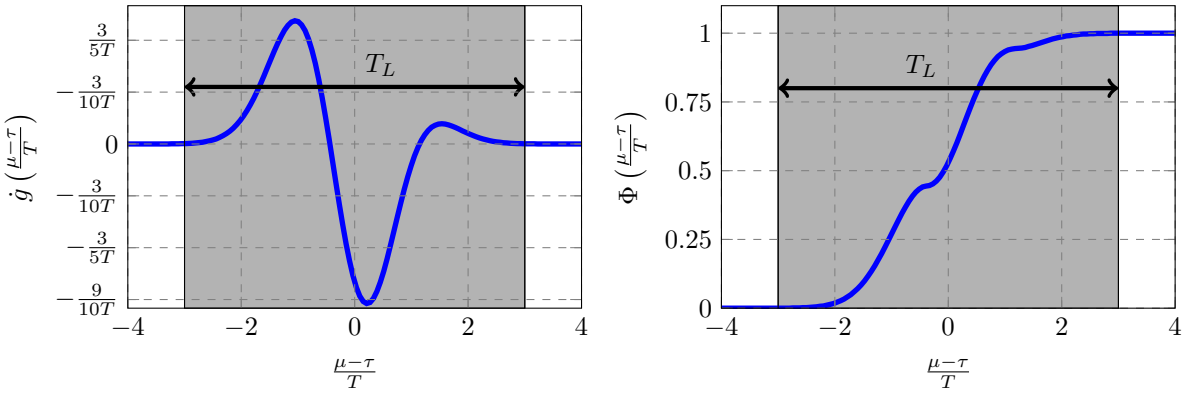


Figure 3.6: Approximation of the first-order derivative of the Hermite diff. $g = g_{N,T,\vartheta,\mu}$ from Example 3.1 as an FIR filter with window length T_L using the distribution function Φ given in (3.9).

with m_0 the moment of $g^{(n)}$ of order 0 defined as

$$\mathrm{m}_0\left(\left|g^{(n)}\right|\right) = \int_{-\infty}^{\infty} \left|g^{(n)}(\tau)\right| \mathrm{d}\tau.$$

To ensure that the truncation at both interval ends is approximately equal it is sufficient to choose $\Phi(\mu - \tau^-) = 1 - \Phi(\mu + \tau^+)$. Assuming that f is continuous, the error stemming from this approximation can be easily upper bounded as

$$\left|\hat{f}^{(n)}(t) - \int_{\mu-\tau^-}^{\mu+\tau^+} g^{(n)}(\tau) f(t-\tau)\mathrm{d}\tau\right| \leqslant 2\mathrm{m}_0\left(\left|g^{(n)}\right|\right) \Phi(\mu-\tau^-) \max_{\tau\in(-\infty,t]} |f(\tau)|\,.$$

Truncating the interval bounds does not only yield an FIR filter for the estimation of the derivatives but also reduces the estimation delay by $\mu - \tau^-$. Using the analytical form of the estimation delay from Table 2.3, the delay after the truncation becomes

$$\delta_t = \tau^- - T\vartheta.$$

Example 3.1 FIR formulation of Hermite diff.

Consider the estimation of the first-order derivative of a function using a Hermite diff. $g_{N,T,\vartheta,\mu}$ with $N = 1$ and ϑ the largest zero of the Hermite polynomial of degree 2. Fig. 3.6 shows the evolutions of the first-order derivative of $g_{N,T,\vartheta,\mu}$ and the distribution function Φ from (3.9). The resulting FIR filter is given by the area in grey where the window is computed such that it starts when Φ takes values greater than 10^{-5}.

Discrete-time FIR approximation

It is assumed for simplicity that the window length of the diff. is an integral multiple of the sampling period, i.e., $T_L = N_L t_\mathrm{s}$. Then, various discrete-time approximations (see, e.g., (Stroud, 1974, Ch. 3)) of the integral in the form of the sum

$$\hat{f}^{(n)}_{k+\theta} \approx \sum_{i=0}^{L-1} w_i f_{k-i}, \tag{3.10}$$

can be achieved, where θ, L, and w_i depend on the numerical integration method used.

In Kiltz (2017) a simple normalisation factor has been introduced to take into account the error stemming from the discretisation. This has also been recalled, for example, in Othmane et al. (2022). Instead of using the approach in (3.10) the following discretisation scheme shall be considered

$$\hat{f}_{k+\theta}^{(n)} \approx \frac{1}{\Phi} \sum_{i=0}^{L-1} w_i f_{k-i}, \quad \text{with} \quad \Phi = \frac{t_\mathrm{s}^n}{n!} \sum_{k=0}^{L-1} w_k (-k)^n. \tag{3.11}$$

The constant Φ is a normalisation factor that takes into account the error stemming from the discretisation when f is assumed to be a polynomial of degree n. This ensures that the low-frequency amplification of the discrete-time filter corresponds to that of the continuous-time filter.

The tuning and derivation of the diff. have been performed in the continuous-time domain. Thus, the discrete-time diff. in (3.11) should preserve the input-output properties of the filter. The discretisation effects can be assessed by comparing the frequency-domain properties of the continuous-time and discrete-time diffs. Denote by $\mathcal{G}$ the Fourier transform of the continuous-time filter given in (2.46). Let $\mathcal{G}_\mathrm{d}$ be the Fourier transform of the discrete-time filter from (3.11) given as

$$\mathcal{G}_\mathrm{d} : \omega \mapsto \mathcal{G}_\mathrm{d}(\omega) = \frac{1}{\Phi} \sum_{i=0}^{L-1} w_i \mathrm{e}^{-\iota\omega(k+\theta)t_\mathrm{s}}.$$

The effects of the discretisation can then be assessed by considering the error norm

$$\mathcal{J} = \frac{\int_0^\Omega \left|\mathcal{G}_\mathrm{d}(\omega) - (\iota\omega)^n \mathcal{G}(\omega)\right|^2 \mathrm{d}\omega}{\int_0^\Omega \left|(\iota\omega)^n \mathcal{G}(\omega)\right|^2 \mathrm{d}\omega}. \tag{3.12}$$

The frequency Ω has to be chosen according to the frequency interval of interest. A reasonable choice is the Nyquist frequency $\omega_\mathrm{N} = \pi/t_\mathrm{s}$, i.e., $\Omega = \omega_\mathrm{N}$. The latter error norm has been introduced in Kiltz (2017) for the analysis of Jac. diffs. This error norm enables a discretisation analysis that does not depend on the properties of the function f. To interpret $\mathcal{J}$, assume that the filter takes as input a band-limited white noise with bandwidth Ω. Then, $\mathcal{J}$ is the ratio of the amplification of the average noise power stemming from the discretisation error and that stemming from the continuous-time filter. Thus, $\mathcal{J}$ is a measure that may help to decide if the discretisation error can be neglected in comparison to the noise amplification of the continuous-time diff. Note that the error stemming from the truncation of the integration interval is also considered by this error norm.

In the following, three discretisation schemes are discussed: the mid-point rule, the trapezoidal rule, and the zero-order hold rule.

Mid-point rule Using the mid-point integration rule yields

$$\theta = \frac{1}{2}, \quad w_k = t_\mathrm{s} g_{k+\theta}^{(n)}, \quad L = N_L.$$

Thus, the delay is reduced by half a sampling period. The formulation in (3.11) then requires $2N_L - 1$ operations (N_L multiplications and $N_L - 1$ additions) as well as the storage of N_L filter coefficients and past values of f.

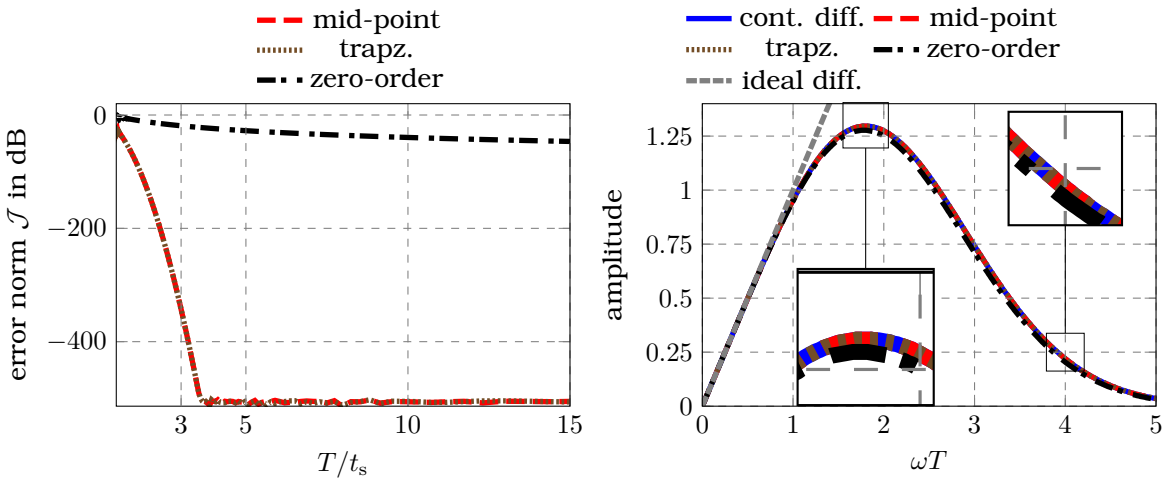


Figure 3.7: Evaluation of the effects of the discretisation on the diff. discussed in Example 3.2: Error norm $\mathcal{J}$ from (3.12) for different discretisation methods and amplitude spectrum of the diff. for $T/t_s = 3$.

Trapezoidal rule Using the trapezoidal integration rule yields

$$\theta = 0, \quad L = N_L + 1, \quad w_k = \begin{cases} t_s g^{(n)}_{k+\theta}/2, & \text{for} \quad k \in \{0, N_L\}, \\ t_s g^{(n)}_{k+\theta}, & \text{otherwise.} \end{cases}$$

This approach requires two more operations than that of the mid-point rule and the storage of one more filter coefficient and past value of f.

Zero-order hold For the first-order hold approach, the function f is assumed to be constant during a sampling interval, i.e., $f(\tau) = f_k$ for $\tau \in (t_{k-1/2}, t_{k+1/2}]$. This yields

$$\hat{f}^{(n)}(t) \approx \sum_{k=0}^{N_L-1} \int_{t_k}^{t_{k+1}} g^{(n)}(\tau) f_k \mathrm{d}\tau.$$

Thus,

$$\theta = \frac{1}{2}, \quad L = N_L, \quad w_k = g^{(n-1)}_{k+1} - g^{(n-1)}_k.$$

for $n > 0$. As for the mid-point rule, the delay is reduced by half a sampling period.

Example 3.2 Discretisation of a Hermite diff. approximated as an FIR filter

Consider the estimation of the first-order derivative of a function using a Hermite diff. $g_{N,T,\vartheta,\mu}$ with $N = 1$ and ϑ the largest zero of the Hermite polynomial of degree 2. Fig. 3.7 shows the evolution of the error norm $\mathcal{J}$ from (3.12) for the three discretisation approaches as a function of T/t_s. While the mid-point and the trapezoidal rules show similar values for $\mathcal{J}$, it can be seen that the zero-order hold discretisation does not preserve well the frequency-domain properties of the diff. This is especially true for small values of T/t_s and can be seen in the right part of the figure where the amplitude spectrum of a diff. for $T/t_s = 3$ is shown.

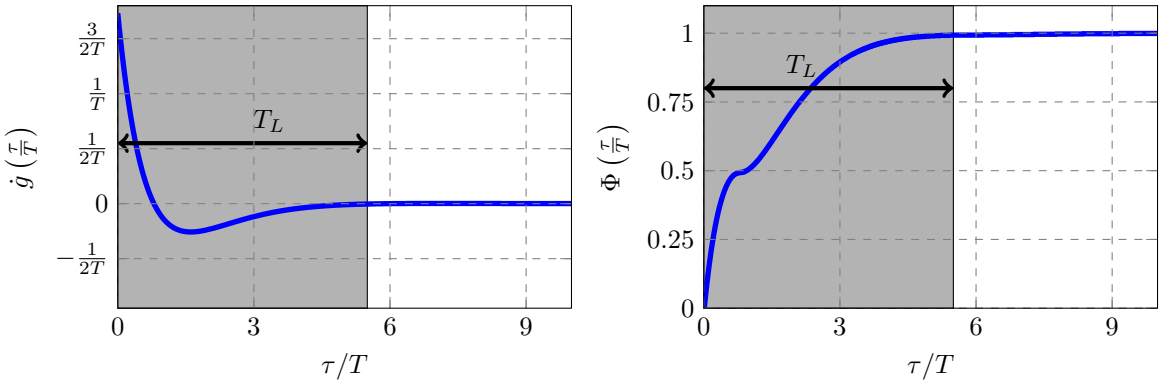


Figure 3.8: Approximation of the first-order derivative of the Lag. diff. $g = g^{(\alpha)}_{N,T,\vartheta}$ from Example 3.3 as an FIR filter with window length T_L using the distribution function Φ given in (3.9).

3.2.2 Implementation of Laguerre differentiators

Recall from Section 2.4.2 that the n-th order derivative of a function f can be approximated as

$$\hat{f}^{(n)}(t) = \int_{-\infty}^{t} g^{(n)}(t-\tau) f(\tau) \mathrm{d}\tau, \quad g = g^{(\alpha)}_{N,T,\vartheta}, \tag{3.13}$$

where $g^{(\alpha)}_{N,T,\vartheta}$ is defined in Definition 2.7.

FIR-Filter approximation

Due to the multiplication of a decaying exponential function with polynomials in $g^{(\alpha)}_{N,T,\vartheta}$ the kernel vanishes at one boundary of the interval. Moreover, for $\alpha \neq 0$ the diff. satisfies $g^{(\alpha)}_{N,T,\vartheta}(0) = 0$. Increasing α flattens the impulse response for small arguments and shifts its maximum towards the right such that $g^{(\alpha)}_{N,T,\vartheta}(t)$ is a small value for $t \to 0$. This can be seen in Figs. 2.7 to 2.9 for different parametrisations. Thus, the approach developed to truncate the integration interval for Hermite diff. in Section 3.2.1 can also be used for Lag. diffs. to approximate (3.13) as FIR filter of the form

$$\hat{f}^{(n)}(t) \approx \int_{t-\tau^-}^{t-\tau^+} g^{(n)}(t-\tau) f(\tau) \mathrm{d}\tau, \quad \tau^- < \tau^+ \geqslant 0. \tag{3.14}$$

The distribution function (3.9) can again be used to compute τ^+ and τ^-. Thus, the approximated derivative can be computed by an FIR filter with window length $T_L = \tau^- - \tau^+$.

For the assessment of the effects of the discretisation on the frequency-domain properties the approaches discussed in Section 3.2.1 for Hermite diffs. can also be used for Lag. diffs.

Example 3.3 FIR approximation of a Lag. diff.

Assume that the first-order derivative of a function $f : t \mapsto f(t)$ has to be approximated using a Lag. diff. with the parametrisation $N = 1$ and $\alpha = 1$. Then, the approximation of the first-order derivative is

$$\hat{\dot{f}}(t) = \int_{-\infty}^{t} \dot{g}(t-\tau) f(\tau) \mathrm{d}\tau, \quad g = g^{(1)}_{1,T,\vartheta}. \tag{3.15}$$

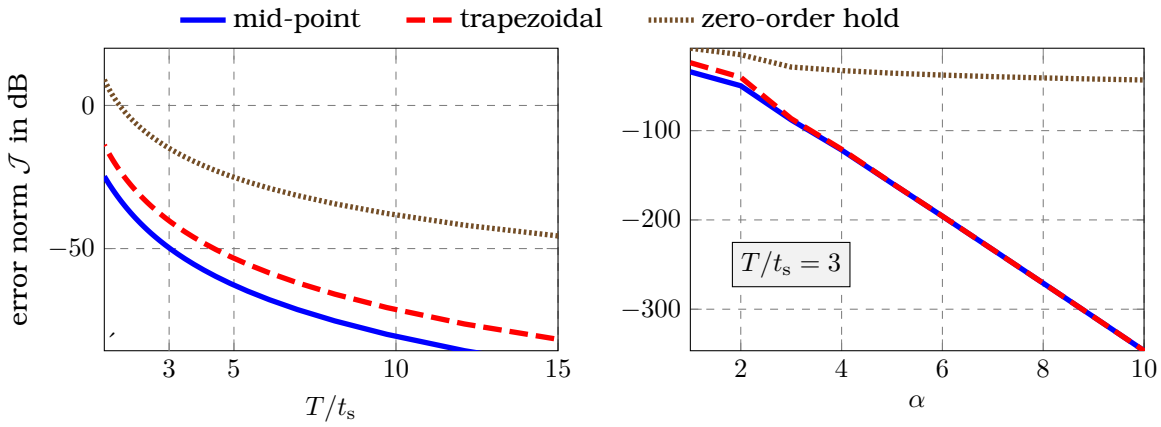


Figure 3.9: Evaluation of the effects of the discretisation of the Lag. diff. discussed in Example 3.4: Error norm $\mathcal{J}$ from (3.12) for different discretisation methods as a function of T/t_s and α.

Fig. 3.8 shows the filter window when the integration interval is truncated as in (3.14). The resulting FIR filter is given by the area in grey where the window is computed such that it starts when Φ takes values greater than 10^{-5}.

Example 3.4 Discretisation analysis of a Lag. diff.

Consider the Lag. diff. from Example 3.3. Fig. 3.9 shows the variation of the error norm $\mathcal{J}$ as a function of T/t_s and α. As for the Hermite diff. in Example 3.2, the mid-point and trapezoidal rules preserve better the frequency domain characteristics of the diff. than the zero-order hold method. Furthermore, increasing α decreases $\mathcal{J}$.

IIR-Filter implementation

Consider the following example, where a Lag. diff. is used for the approximation of the first derivative of a function. This simple example shows that a solution of a specific differential equation exponentially converges towards the approximated derivative in (3.13). The observations are then generalised and the discrete-time implementation is discussed. It is assumed in the sequel that $\alpha \in \mathbb{N}$.

Example 3.5 Estimation of first-order derivative

Consider the linear time-invariant (LTI) system

$$\left(T\frac{\mathrm{d}}{\mathrm{d}t}+1\right)^3 y = \frac{\mathrm{d}}{\mathrm{d}t}\left(-1+T(\vartheta-3)\frac{\mathrm{d}}{\mathrm{d}t}\right) f. \tag{3.16}$$

with input f and output y for some arbitrary initial conditions. Its transfer function is

$$h_f^y(\imath\omega) = \frac{\imath\omega\left(-1+T\left(\vartheta-3\right)\imath\omega\right)}{\left(T\imath\omega+1\right)^3}$$

and corresponds to the Fourier transform of the Lag. diff. considered in Example 3.3. It is then easy to show that the solution of (3.16) converges towards (3.15) for arbitrary initial conditions since the system possesses the eigenvalue $-1/T$ with the algebraic multiplicity of three. If the initial conditions are appropriately chosen, the solution even corresponds to the latter approximation of the first-order derivative in (3.15). The differential equation can then be discretised and implemented as a infinite-impulse response (IIR) discrete filter.

Continuous-time formulation Consider the stable LTI system

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+T^{-1}\right)^{\bar{\alpha}} y = \left(\sum_{m=0}^{N} \frac{h^{(\alpha)}_{m,N,\vartheta}}{T^{m+\alpha+1}} \frac{\mathrm{d}}{\mathrm{d}t}^{N+n-m}\right) f, \quad \bar{\alpha} = \alpha + N + 1, \tag{3.17}$$

for some arbitrary initial conditions with

$$h^{(\alpha)}_{m,N,\vartheta} = \sum_{k=0}^{N-m} \binom{N-k}{m} \mathrm{L}^{(\alpha)}_k(\vartheta) \tag{3.18}$$

where f is the input and y the output. Its transfer function is

$$h^{y}_{f}(\iota\omega) = \frac{1}{T^{\alpha+1}} \sum_{m=0}^{N} \mathrm{L}^{(\alpha)}_m(\vartheta) \frac{(\iota\omega)^{m+n}}{(\iota\omega + T^{-1})^{\alpha+m+1}}$$

which is the Fourier transform of $\left(g^{(\alpha)}_{N,T,\vartheta}\right)^{(n)}$.

The solution of (3.17) exponentially converges to the approximation of the n-th order derivative of f defined in (3.13) for arbitrary initial conditions. While the error stemming from the initial conditions tends to zero for small T, the transient behaviour needs to be analysed further. In fact, it is straightforward to show that it may contain terms of the form $\mathrm{e}^{-\frac{t}{T}} T^{-\rho}$, for some ρ, which approach an impulse function as T tends to zero. Thus, the transient phase may contain a peak with amplitude of order $1/T$. This phenomenon is known as peaking phenomenon and has been already observed for HG observers in Esfandiari and Khalil (1992), for example. The relation to these observers is analysed in detail in the next section. To conclude, the initial conditions can be specified arbitrarily. However, care must be paid to the behaviour in the transient phase.

Discrete-time formulation For the computation of y in (3.17) in the presence of discrete measurements the differential equation has to be discretised. It can be approximated by a difference equation of the form

$$P(\delta)\, \xi_k = Q(\delta)\, f_k \tag{3.19}$$

with P and Q polynomials and δ the backward shift operator. The output y in (3.17) is then approximated by $y(kt_\mathrm{s}) = \xi_k$. The polynomials P and Q depend on the choice of the discretisation approach and can easily be obtained using the approaches found in (Franklin, Powell, Workman, et al., 1998, Ch. 6) or (Oppenheim & Schafer, 1975, Ch. 5). The bilinear[2] discretisation scheme, for example, preserves the stability properties of the continuous-time filter. The degrees r_P and r_Q of P and Q satisfy $r_P = \bar{\alpha}$ and $r_Q \leqslant \bar{\alpha}$. Thus, implementing (3.19)

[2]The bilinear discretisation scheme is also known as Tustin's method after the British engineer Arnold Tustin 1899-1994.

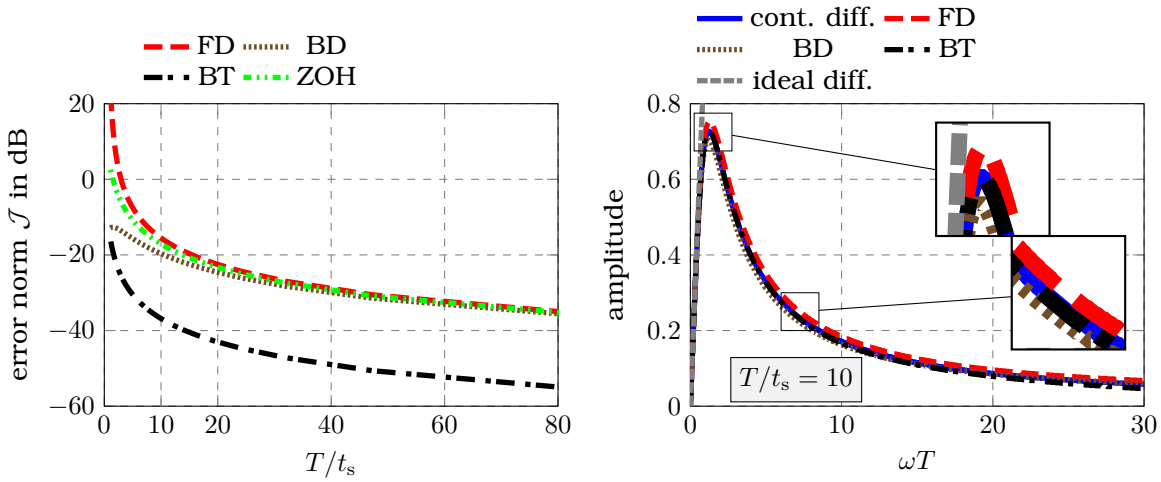


Figure 3.10: Evaluation of the effects of the discretisation on the diff. discussed in Example 3.6: Error norm $\mathcal{J}$ from (3.12) for different discretisation methods and the amplitude spectrum of the diff. for $T/t_s = 10$.

using the direct form I realisation[3] (see, e.g., (Oppenheim & Schafer, 1975, Sec. 4.3)) requires storing r_P and r_Q past values of ξ and f, respectively. The total number of filter coefficients to be stored is $r_P + r_Q + 1$. The number of required operations is $2(r_P + r_Q) + 1$ (more specifically $r_P + r_Q + 1$ multiplications and $r_P + r_Q$ additions). Thus, the number of required operation (multiplication, addition, and delays) depends linearly on N and α. It does not depend on T, the parameter influencing the cutoff frequency as opposed to the FIR implementation.

As for the approach developed for the FIR approximation, the effect of the discretisation can be assessed by the error norm in (3.12) by computing the discrete transfer function corresponding to the difference equation in (3.19). The following example considers the discretisation of the LTI system discussed in Example 3.5.

Example 3.6 Discretisation of an IIR Lag. diff.

Consider again the diff. from Example 3.5. Fig. 3.10 shows the evolution of the error norm in (3.12) assessing the effect of the discretisation on the frequency domain properties of the diff. for different discretisation schemes. The bilinear transform (BT) shows the lowest error. The forward difference (FD) yields big errors for small T/t_s. This is a known problem since this approach does not preserve the stability as discussed in Franklin et al. (1998); Oppenheim and Schafer (1975), for example.

The difference equation of the form (3.19) for the BT is

$$\left(p_3\delta^3 + p_2\delta^2 + p_1\delta + p_0\right)\xi_k = \left(q_3\delta^3 + q_2\delta^2 + q_1\delta + q_0\right) f_k$$

[3]The direct form I is not a realisation with the minimum number of shift operators as discussed in (Oppenheim & Schafer, 1975, Sec. 4.3). Different other forms known as canonic forms can be derived and use the minimum number of shift operators. However, the direct form I is used in this section due to its simplicity.

with

$$\begin{aligned} p_0 &= 8\,T^3 + 12\,t_\mathrm{s}\,T^2 + 6\,t_\mathrm{s}^2 T + t_\mathrm{s}^3, & q_0 &= -4\,T t_\mathrm{s}\,\vartheta + 12\,t_\mathrm{s}\,T + 2\,{t_\mathrm{s}}^2, \\ p_1 &= -24\,T^3 - 12\,t_\mathrm{s}\,T^2 + 6\,t_\mathrm{s}^2 T + 3\,t_\mathrm{s}^3, & q_1 &= 4\,T t_\mathrm{s}\,\vartheta - 12\,t_\mathrm{s}\,T + 2\,{t_\mathrm{s}}^2, \\ p_2 &= 24\,T^3 - 12\,t_\mathrm{s}\,T^2 - 6\,t_\mathrm{s}^2 T + 3\,t_\mathrm{s}^3, & q_2 &= 4\,T t_\mathrm{s}\,\vartheta - 12\,t_\mathrm{s}\,T - 2\,{t_\mathrm{s}}^2, \\ p_3 &= -8\,T^3 + 12\,t_\mathrm{s}\,T^2 - 6\,t_\mathrm{s}^2 T + t_\mathrm{s}^3, & q_3 &= -4\,T t_\mathrm{s}\,\vartheta + 12\,t_\mathrm{s}\,T - 2\,{t_\mathrm{s}}^2. \end{aligned}$$

The latest ξ_k can then be computed as

$$\xi_k = \left(\bar{p}_3\delta^3 + \bar{p}_2\delta^2 + \bar{p}_1\delta\right)\xi_k + \left(\bar{q}_3\delta^3 + \bar{q}_2\delta^2 + \bar{q}_1\delta + \bar{q}_0\right) f_k$$

with $\bar{p}_i = -p_i/p_0$ and $\bar{q}_i = q_i/p_0$. Thus, seven filter coefficients and three past values of f and ξ need to be stored. At each instant k the total number of required operations is 13.

The value of ξ_k converges towards the approximation of the first-order derivative of f from a Lag. diff. given in (3.15).

Simultaneous derivative estimation

In numerous applications the estimation of derivatives of several orders is required. Moreover, it is often important that all derivatives are filtered by the same filter (recall from Table 2.4 that the approximated derivative is the output of a filter driven by the sought derivative). This ensures that all approximations have the same delay and the same amplitude and phase distortion. The following analysis also shows that HG diff. are special cases of Lag. diff.

Continuous-time formulation The simultaneous approximation of derivatives of orders up to n can be easily performed using the discussion from the previous section. The approximation of all derivatives of orders $m \in \{0, \dots, n\}$ is considered. Then, the parameter α of the Lag. diffs. has to satisfy $\alpha > n - 1$ as given in Definition 2.7. Following the lines of Section 3.2.2, it follows that the solution of the stable LTI system

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + T^{-1}\right)^{\bar{\alpha}} y = \left(\sum_{m=0}^{N} \frac{h^{(\alpha)}_{m,N,\vartheta}}{T^{m+\alpha+1}} \frac{\mathrm{d}}{\mathrm{d}t}^{N-m}\right) f, \quad \bar{\alpha} = \alpha + N + 1, \tag{3.20}$$

with $h^{(\alpha)}_{m,N,\vartheta}$ defined in (3.18), converges exponentially to the approximation

$$\hat{f}(t) = \int_{-\infty}^{t} g^{(\alpha)}_{N,T,\vartheta}(t-\tau) f(\tau) \mathrm{d}\tau.$$

Consider the minimal state realisation (see, e.g., (Kailath, 1980, Ch. 2&6)) for the system (3.20) given by

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\boldsymbol{x} + \boldsymbol{b}f, \\ y &= \boldsymbol{c}^T\boldsymbol{x}, \end{aligned} \tag{3.21a}$$

with $\boldsymbol{x}(t) \in \mathbb{R}^{\bar{\alpha}}$, suitable initial condition $\boldsymbol{x}(t_0)$, and

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \\ -\frac{a_{0,\bar{\alpha}}}{T^{\bar{\alpha}}} & -\frac{a_{1,\bar{\alpha}}}{T^{\bar{\alpha}-1}} & \dots & \dots & -\frac{a_{\bar{\alpha}-2,\bar{\alpha}}}{T^2} & -\frac{a_{\bar{\alpha}-1,\bar{\alpha}}}{T} \end{bmatrix}, \; b = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \; c = \begin{bmatrix} \frac{h^{(\alpha)}_{N,N,\vartheta}}{T^{\bar{\alpha}}} \\ \frac{h^{(\alpha)}_{N-1,N,\vartheta}}{T^{\bar{\alpha}-1}} \\ \vdots \\ \frac{h^{(\alpha)}_{0,N,\vartheta}}{T^{\bar{\alpha}-N}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.21b}$$

where $a_{m,\bar{\alpha}} = \binom{\bar{\alpha}}{m} = \bar{\alpha}!/(m!(\bar{\alpha}-m)!)$ and $h^{(\alpha)}_{m,N,\vartheta}$ is defined in (3.18).

The chain of integrators in (3.21) yields that the linear combination

$$y_l = \sum_{m=0}^{N} \frac{h^{(\alpha)}_{N-n,N,\vartheta}}{T^{\bar{\alpha}-m}} x_{m+l}, \quad l \in \{1, \dots, \alpha\}, \tag{3.22a}$$

converges exponentially to the approximation of the l-th order derivative computed using a Lag. diff. $g^{(\alpha)}_{N,T,\vartheta}$ of order α and given as

$$\hat{f}^{(l)} = \int_{-\infty}^{t} g^{(l)}(t-\tau) f(\tau) \mathrm{d}\tau, \quad g = g^{(\alpha)}_{N,T,\vartheta}. \tag{3.22b}$$

It is important to observe here that all the derivatives $f^{(l)}$ act as inputs to the same filter and thus the approximations are a set of matched time derivatives. An important property is that all derivatives have the same estimation delay. State realisations of the form (3.21) with outputs (3.22a) are known as state variable filters (see, e.g., Hofmann et al. (1966); Peter and Isermann (1990); Young (1981)). They have been widely used in the context of parameter estimation for linear systems (see, e.g., Isermann (1977, 2013b); Young (1981)).

In addition to the approximations in (3.22b) it also follows that

$$\bar{y}_l = \sum_{m=0}^{N-l} \frac{h^{(\alpha)}_{N-n,N,\vartheta}}{T^{\bar{\alpha}-m}} x_{\alpha+1+l+m}, \quad l \in \{0, \dots, N\}, \tag{3.23a}$$

exponentially converges to

$$\hat{f}^{(\alpha+1+l)} = \int_{-\infty}^{t} g^{(\alpha+1+l)}(t-\tau) f(\tau) \mathrm{d}\tau, \quad g = g^{(l)}_{N+\alpha-l,\vartheta,T}. \tag{3.23b}$$

This can be verified by computing the input-output relationship between the input f and the output $\bar{y}_l$. These approximations, however, do not have the same properties as those in (3.22b). In particular, the estimation delays are different.

Due to the importance of this observation, the result is summarised in the following corollary.

Corollary 3.1: Simultaneous approximation of derivatives using Lag. diff.

Let f be a function that is bounded and differentiable N times with bounded derivatives up to the N-th order and piecewise continuous N-th order derivative. Consider the system (3.21). Then, the linear combinations (3.22a) and (3.23a) converge to the approximations of Lag. diffs. in (3.22b) and (3.23b), respectively.

The error stemming from the initial conditions converges exponentially to zero. The transient phase, however, may show significant overshoots. In fact, due to the eigenvalue $-1/T$ of algebraic multiplicity $N+\alpha+1$ the solution contains expressions of the form $\mathrm{e}^{-\frac{t}{T}}T^{-\rho}$, for some ρ. These expressions approach an impulse function as T tends to zero. Thus, when these approaches are used in closed-loop control care must be paid. The problem can, however, be easily solved by saturating the approximated derivative outside a compact region of interest as for HG observers (see, e.g., Esfandiari and Khalil (1992)).

Relation to high-gain differentiators The appearance in (3.21) of negative powers of the parameter T that scales the orthogonality interval of the Laguerre polynomials raises the question whether there exists a relation to the well known HG diffs. from Dabroom and Khalil (1997, 1999); Khalil (2009, 2017); Vasiljevic and Khalil (2006, 2008).

Let f be a function that is bounded and differentiable n times with bounded derivatives up to the n-th order and piecewise continuous n-th order derivative. Consider the HG diff. of order n with input f, given as

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{b}f, \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \in \mathbb{R}^n \tag{3.24a}$$

with

$$A = \begin{bmatrix} -c_0/T & 1 & 0 & 0 & \dots & 0 \\ -c_1/T^2 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -c_{n-3}/T^{n-2} & 0 & \dots & 0 & 1 & 0 \\ -c_{n-2}/T^{n-1} & 0 & \dots & \dots & 0 & 1 \\ -c_{n-1}/T^n & 0 & \dots & \dots & \dots & 0 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} c_1/T \\ c_2/T^2 \\ \vdots \\ c_n/T^n \end{bmatrix}, \tag{3.24b}$$

where the parameters c_l, $l \in \{0, \dots, n-1\}$, are chosen such that the polynomial

$$\lambda \mapsto \lambda^n + \sum_{l=0}^{n-1} c_l \lambda^{n-l} \tag{3.24c}$$

is Hurwitz. The k-th, $k \in \{1, \dots, n\}$, component of $\boldsymbol{x}$ is an approximation of the k-th order derivative of f. As shown in Esfandiari and Khalil (1992), the estimation error decays to $\mathcal{O}(T)$ values after a short transient period of the form $[0, T_0(T)]$, where $\lim_{T\to 0} T_0(T) = 0$.

The following result shows that for a special choice of the roots of the polynomial (3.24c) the solution of the dynamic system (3.24) converges to the approximated derivative from a Lag. diff.

Corollary 3.2: Relation between HG and Lag. diffs.

Let $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ be a function that is bounded and differentiable n times with bounded derivatives up to the n-th order and piecewise continuous n-th order derivative. Consider the system (3.24) and assume that the polynomial in (3.24c) has the root -1 with an algebraic multiplicity of n. Then, there exists positive scalars κ_1 and κ_2 such that the m-th component x_m of the state $\boldsymbol{x}$ satisfies

$$\left| x_m(t) - \int_0^t g_m^{(m-1)}(t-\tau) f(\tau) \mathrm{d}\tau \right| < \kappa_1 \mathrm{e}^{-\kappa_2 t}, \quad g_m(\tau) = g_{n-m,T,0}^{(m-1)}(\tau) \tag{3.25}$$

for all t, $m \in \{1, \ldots, n\}$, and arbitrary initial conditions, where $g_{n-m,T,0}^{(m-1)}$ is a Lag. diff. of order $m-1$ defined in Definition 2.7.

Proof. It can be verified using elementary calculus that

$$\boldsymbol{x}_{\mathrm{p}}(t) = \begin{bmatrix} \int_{t_0}^t g_1(t-\tau) f(\tau) \mathrm{d}\tau \\ \int_{t_0}^t \dot{g}_2(t-\tau) f(\tau) \mathrm{d}\tau \\ \vdots \\ \int_{t_0}^t g_n^{(n-1)}(t-\tau) f(\tau) \mathrm{d}\tau \end{bmatrix}$$

is a particular solution of (3.24). Then, the general solution of (3.24) is

$$\boldsymbol{x}(t) = \mathrm{e}^{A(t-t_0)} \boldsymbol{x}_0 + \boldsymbol{x}_{\mathrm{p}}(t),$$

where each component x_m, $m \in \{1, \ldots, n\}$, of $\boldsymbol{x}$ satisfies

$$x_m(t) = \mathrm{e}^{-t/T} p_m(t) + \int_{t_0}^t g_m(t-\tau) f(\tau) \mathrm{d}\tau,$$

for some polynomials p_m. Recalling the definition of $g_{n-m,T,0}^{(m-1)}(\tau)$ from (2.49) yields the result (3.25). □

Corollary 3.2 shows that the state components x_m, $m \in \{1, \ldots, n-1\}$, tend towards the approximation stemming from a delay-free Lag. diff. of order $m-1$. The last component tends toward the approximation stemming from a minimal Lag. diff. of order $n-1$ and is thus delayed. This component corresponds to the approximation of the $(n-1)$-th order derivative of f. Following Table 2.3 and Corollary 2.3 the delay is $\delta_t = Tn$.

Discrete-time formulation In the presence of discrete measurements the differential equation (3.21) has to be discretised. It can be approximated by a discrete-time system

$$\begin{aligned} \boldsymbol{\zeta}_{k+1} &= A_{\mathrm{d}} \boldsymbol{\zeta}_k + \boldsymbol{b}_{\mathrm{d}} f_k, \\ \xi_k &= \boldsymbol{c}_{\mathrm{d}}^{\mathrm{T}} \boldsymbol{\zeta}_k + d_{\mathrm{d}} f_k \end{aligned} \tag{3.26}$$

with $k \in \{0, 1, \ldots\}$ and the initial conditions $\boldsymbol{\zeta}_{t_0} = A_{\mathrm{d},0} \boldsymbol{x}(t_0) + \boldsymbol{b}_{\mathrm{d},0} u(t_0)$. Then, $y(kt_{\mathrm{s}}) \approx \xi_k$. The involved matrices and vectors depend on the discretisation scheme (see, e.g., Franklin et al. (1998); Oppenheim and Schafer (1975, 1989); Stroud (1974)).

3.2.3 Implementation of Jacobi differentiators

Recall from Section 2.4.3 that the n-th order derivative of a function f can be approximated as

$$\hat{f}^{(n)}(t) = \int_{t-T}^{t} g^{(n)}(t-\tau) f(\tau) \mathrm{d}\tau, \quad g = g_{N,T,\vartheta}^{(\alpha,\beta)}, \tag{3.27}$$

where $g_{N,T,\vartheta}^{(\alpha,\beta)}$ is defined in (2.67).

Jac. diffs. have initially been discretised in Liu (2011); Mboup et al. (2007); Mboup et al. (2009) using the trapezoidal rule. This yields discrete filters as discussed for Hermite diff. in Section 3.2.1. Further discretisation approaches as the mid-point rule, zero-order and first-order hold have then been discussed in Kiltz and Rudolph (2013). For all approaches the approximation (3.27) becomes

$$\hat{f}_{k+\theta}^{(n)} = \sum_{i=0}^{L-1} w_k f_{i-k}, \tag{3.28}$$

where w_k is the discrete filter coefficients and the parameter θ depends on the discretisation approach (see, e.g., the detailed discussion for the discretisation of Hermite diff. in Section 3.2.1). This convolution sum has been evaluated non-recursively in each sampling step in all previously cited publications. A drawback of this approach is that the number of required arithmetic operations per step and the number of filter coefficients to be stored increase linearly with increasing filter window length and thus decreasing cutoff frequency. Thus, they may become very large.

A procedure for the recursive evaluation of the Jacobi derivative estimators in the context of parameter estimation has been discussed in Gensior (2009); Gensior, Weber, Güldner, and Rudolph (2007); Gensior et al. (2008). In this procedure, the number of computational operations per step and the number of filter coefficients to be stored become independent of the filter window length. This implementation can significantly improve the computational burden and storage efficiency of the discretised diffs. However, this implementation procedure is not applicable for all parameter combinations of the Jac. diffs., and the accuracy of the discrete-time approximation deteriorates as the stopband slope increases. These issues have motivated the investigation of more complex recursive implementation procedures in Kiltz (2017). The mid-point discretisation rule plays an important role, as can be seen in the following motivating example.

Example 3.7 Example recursive implementation of Jac. diff. from Kiltz (2017)

Consider the implementation of a Jac. diff. for the estimation of the first-order derivative of a function f with $N = 0$ and $\alpha = \beta = 1$ from (Kiltz, 2017, Sec. 2.4). The mid-point discretisation rule yields the approximated derivative

$$\dot{\hat{f}}_{k+\frac{1}{2}} = \sum_{i=0}^{L-1} w_i f_{k-i}, \quad w_i = \frac{6(L-1-2i)}{L^2 T},$$

at step k with the filter window length satisfying $T = Lt_\mathrm{s}$. Computing this approximation at each step requires $2L-1$ operations. Moreover, L filter coefficients and past values of f have to be stored. The latter discrete convolution can however be implemented as a

recursion. Then, the approximation is $\hat{\dot{f}}_{k+\frac{1}{2}} = \frac{6}{(L^2-1)T} z_k$, with

$$z_k = d_1(f_k - f_{k-L-1}) - d_2(f_{k-1} - f_{k-L}) + 2z_{k-1} - z_{k-2}, \quad d_1 = L-1, \quad d_2 = L+1.$$

To ensure the stability of this recursion the implementation needs to be performed using fixed point arithmetic. This can be achieved, for example, when f is the output of an analog-to-digital converter. This implementation is independent of the filter window length and requirs only nine operations at each step. However, it requires the history of f in the entire window.

The ideas shown in Example 3.7 have been generalised in Kiltz (2017) for arbitrary values of N, α, β, and ϑ. It is shown in (Kiltz, 2017, Sec. 3.4.4) that the output y of a Jac. diff. can be computed as

$$y_k = -\sum_{m=0}^{r+1} \binom{r+1}{m} (-1)^m y_{k-m} + \frac{1}{\Theta} \sum_{i=0}^{q} \sum_{m=0}^{r} c_{i,m} f_{k-k_i}$$

with Θ the normalisation constant from Section 3.2.1. The quantities r, q, $c_{i,m}$ and k_i are parameters that depend on the Jac. diff. that can be taken from (Kiltz, 2017, Eq. (B.79)). The parameter r and q are independent of T and thus of the filter window length. For $\alpha, \beta \in \mathbb{N}$, for example, $q = 1$ and $r = \alpha + \beta + N - n$.

For the stable implementation of the latter recursion a fixed point arithmetic is required. To ensure that all coefficients $c_{i,m}$ are rational numbers, an approximation of the discretised filter coefficients using discrete Chebyshev polynomials has been used. The effect of this approximation and the required quantisation on the filter properties has been discussed in detail by introducing an error norm that compare these effects with those stemming from the discretisation of the convolution integral in (3.27). The reader is referred to (Kiltz, 2017, Sec. 3.4) for further details.

Evaluating the approximation using the discrete convolution in (3.28) requires $2L-1$ arithmetic operations in each step and the storage of L filter coefficients and past values of the measurement, respectively. These numbers clearly increase linearly with increasing L. Contrarily, the recursive implementation requires only $(q+1)(2r+1)+2$ operations which is independent of L. The approach requires storing $L+r$ past measurements, $r+1$ past outputs, and $(q+2)(r+1)$ filter coefficients. Thus, the total required storage capacities are $(q+4)(r+1)+L-1$ elements compared to $2L$ for the sum in (3.28). Thus, for large enough L the recursive implementation significantly reduces the computational and storage requirements.

A major drawback of this implementation approach is the importance of the initialisation of the recursion and the required fixed point arithmetic. For the estimation of the third-order derivative using a diff. with $\alpha = \beta = 4$, $N = 0$, and $L = 200$ a representation with at least 33 bits is required.

3.3 Application of systematic tuning guidelines

The systematic tuning guidelines addressed in the last section shall be applied in two case studies. First, the estimation of the first and second-order derivatives of a disturbed experimental signal are approximated. The approximation results of the three diffs. are compared with respect to the quality, computational burden, and memory requirements. Then, the ap-

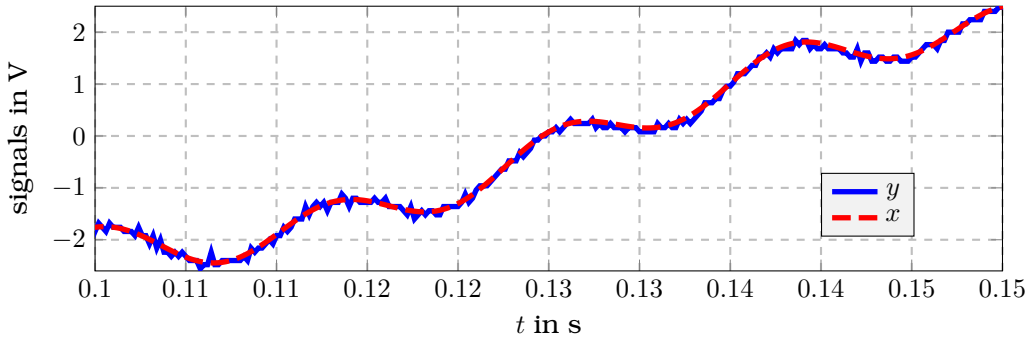


Figure 3.11: True and measured voltage signals from Example 3.8.

proximate inversion of analogue anti-aliasing filters is investigated. Convincing experimental results are presented.

3.3.1 Derivative approximation

The tuning approaches proposed in Section 3.1 are demonstrated using experimental data from the following example.

Example 3.8 Derivative approximation using a laboratory measurement

The voltage applied to an electric circuit by a signal generator is measured using an oscilloscope. The applied voltage is

$$x(t) = a_1 \sin(2\pi f_1 t) + a_2 \sin(2\pi f_2 t), \quad f_1 = 10\,\text{Hz},\ f_2 = 80\,\text{Hz},\ a_1 = 2.15\,\text{V},\ a_2 = 0.4\,\text{V},$$

and the disturbed measured voltage is denoted by y. The disturbance is assumed to be additive. Fig. 3.11 shows the time evolution of x and y. The signal to noise ratio (SNR) is

$$\text{SNR} = 10 \log \left(\frac{\sum_{k=0}^{N_y} y_k^2}{\sum_{k=0}^{N_y} (y_k - x_k)^2} \right) = 27\,\text{dB},$$

with N_y the number of samples of the signal. The total length of the considered signal is $0.4\,\text{s}$. The approximation of the first-order derivative of x using y shall be considered. The measurement is performed at a frequency of $5\,\text{kHz}$.

The systematic tuning approach is used to find filter parameters to approximate the first and second-order derivative of x using y from Example 3.8. Hermite, Laguerre, and Jacobi diffs. are considered. In the following, the approximation of the n-th order derivative of x computed using y is denoted by $\hat{y}^{(n)}$. Similarly, the approximation computed using x is denoted by $\hat{x}^{(n)}$.

To perform a quantitative comparison of the approximation results the output signal to noise ratio (SNRout) of the approximation defined as

$$\text{SNRout} = \frac{\sum_{k=T_L}^{N_s} \left| \hat{x}_k^{(n)} \right|^2}{\sum_{k=T_L}^{N_s} \left| \hat{x}_k^{(n)} - \hat{y}_k^{(n)} \right|^2}, \quad n \in \{0, \ldots, 2\},$$

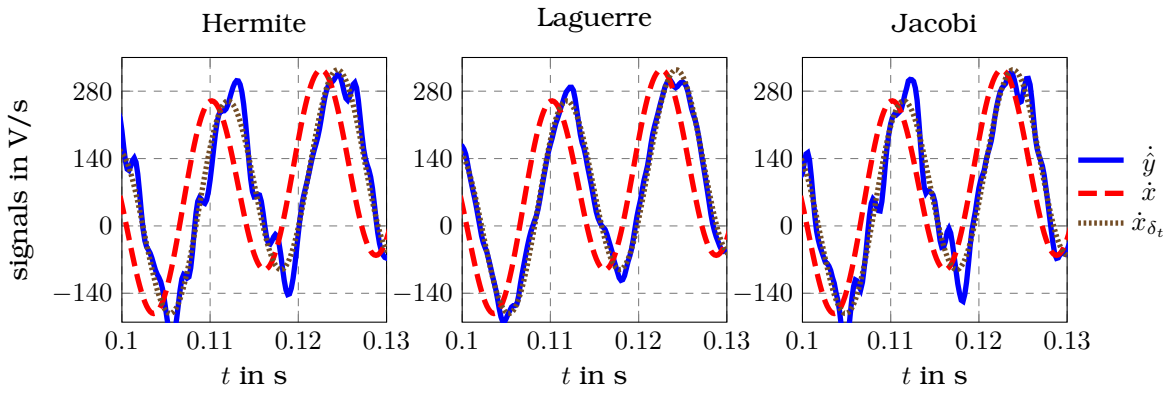


Figure 3.12: Approximation results for the first-order derivative of the signal x from Example 3.8. The parameters of the diffs. are summarised in Table 3.1.

and the signal to bias ratio (SBR) given as

$$\mathrm{SBR} = \frac{\sum_{k=T_L}^{N_s} \left| x_{\delta_t,k}^{(n)} \right|^2}{\sum_{k=T_L}^{N_s} \left| x_{\delta_t,k}^{(n)} - \hat{x}_k^{(n)} \right|^2}, \quad n \in \{0, \ldots, 2\},$$

are considered, where $x_{\delta_t}^{(n)}$ is the n-th order derivative of x delayed by the delay δ_t of the differentiator. While the SNRout takes into account the error stemming from the measurement disturbance, the SBR describes the error stemming from the generalised Fourier series truncation. The parameter T_L denotes the filter window length.

For all diffs. the parameter N is equal to 1. According to Corollary 2.2 the obtained diffs. in a parametrisation for a delayed approximation have a degree of exactness (DoE) equal to two and three for the approximation of the first-order and second-order derivatives, respectively. The resulting delay can be computed according to Corollary 2.3. Contrarily, when a delay-free approximation is considered the DoEs are equal to one and two, respectively. All diffs. are discretised using the midpoint rule and implemented as FIR filters as discussed in Section 3.2. The windows of the filters are computed such that the window starts when the distribution function from (3.9) takes values greater than 10^{-4}. For all diffs. considered in the sequel the error norm in (3.12) is less than $-80\,\mathrm{dB}$. Thus, the discrete filters preserve the frequency-domain properties of the continuous ones.

Approximation of the first-order derivative

Comparison of approximation results Parametrisations that yield delayed approximations are considered first. Start with a parametrisation where the cutoff frequencies of the Lag. diff. and Jac. diff. are chosen as $\omega_\mathrm{c} = \omega_0 = 8\pi f_2 \mathrm{rad}$. Furthermore, $\alpha = 3$ for the Lag. diff. and $\alpha = \beta = 3$ for the Jac. diff. For the Hermite diff. the parameter T is computed such that the amplitude of the Fourier transform at ω_c is equal to $-3\,\mathrm{dB}$. Table 3.1 summarises the numerical values of the remaining parameters that follow for these parametrisations.

The evolution in time of the approximated derivative using the latter parametrisations of the diffs. is given in Fig. 3.12. It can be seen that the approximation from the Lag. diff. is better than that of the remaining two diffs. This is also confirmed by the SNRout given in Table 3.1 which is equal to $20.0\,\mathrm{dB}$ for the Lag. diff. as opposed to $14.47\,\mathrm{dB}$ and $13.53\,\mathrm{dB}$

Table 3.1: Approximation results for the first-order derivative of the signals from Example 3.8 for $N = 1$, $\omega_c = 8\pi f_2 \mathrm{rad}$. All diffs. are parametrised such that ϑ is a zero of the orthogonal polynomial (OP) of interest with degree $N + 1$. The zero is chosen such that the delay is minimised. The delay and the FIR window length are given as multiples of the sampling period t_s.

	Hermite diff.	Lag. diff.	Jac. diff.
	$\mu = 90\,\mathrm{ms}$, $T = 0.91\,\mathrm{ms}$	$\alpha = 3$, $T = 0.6\,\mathrm{ms}$	$\alpha = \beta = 3$, $T = 3.6\,\mathrm{ms}$
delay	9	8	6
window length T_L	28	55	17
SNRout in dB	14.47	20.0	13.53
SBR in dB	34.84	24.82	44.08

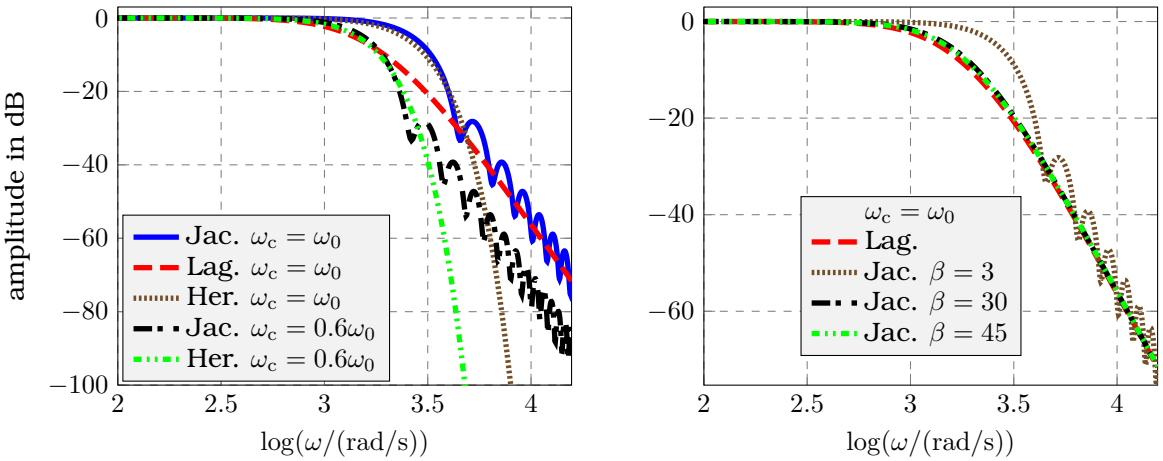


Figure 3.13: Amplitude spectra for the diffs. used for the approximation of the first-order derivative of the signal x from Example 3.8. The parameters of the diffs. are summarised in Table 3.1.

for the Hermite and the Jacobi diffs., respectively. However, the SBR of the Jac. diff. is the highest with $44.08\,\mathrm{dB}$ as opposed to $34.84\,\mathrm{dB}$ and $24.82\,\mathrm{dB}$ for the Hermite and the Laguerre diffs., respectively. These differences can be explained by the differences in the amplitude spectra given in Fig. 3.13. While the Jacobi and Laguerre diffs. have the same filter order, the amplitude evaluated at the cutoff frequency is higher for the former compared to that of the latter. Similarly, for the Hermite diff., the amplitude converges faster to zero for high frequencies than for the remaining two. However, the value at the cutoff frequency is comparable to that of the of the Jac. diff. Moreover, the estimation delay of the Jac. diff. equal to six sampling periods is less than the delays of the Laguerre and the Hermite diffs., which are equal to nine and eight, respectively.

Varying the parameters of the Jac. diff. may yield better results. Fig. 3.14 and Table 3.2 show the results when α and the cutoff frequency are kept identical and β is increased. It can be seen that the SNRout increases when β is increased. For $\beta = 3$ it is equal to $14.47\,\mathrm{dB}$ as opposed to $19.81\,\mathrm{dB}$ for $\beta = 45$. Contrarily, the SBR decreases from $44.08\,\mathrm{dB}$ to $23.20\,\mathrm{dB}$. These findings can be explained by the variation of the amplitude spectrum as a function of β as shown in Fig. 3.13. It is clear that the value of the amplitude of the Fourier transform

Table 3.2: Approximation results for the first-order derivative of the signals from Example 3.8 using a Jac. diff. with $N = 1$, $\alpha = 3$, and a cutoff frequency $\omega_c = \omega_0 = 8\pi f_2 \mathrm{rad}$ and different values of β. The delay and the FIR window length are given as multiples of the sampling period t_s.

β	3	15	30
delay	6	7	8
window length	17	55	60
SNRout in dB	14.47	19.03	19.81
SBR in dB	44.08	33.13	23.20

Table 3.3: Approximation results for the first-order derivative of the signals from Example 3.8 using Jacobi and Hermite diffs. with different values for the cutoff frequency and $N = 1$. For the Jac. diff. $\alpha = \beta = 3$. The delay and the FIR window length are given as multiples of the sampling period t_s and $\omega_0 = 8\pi f_2 \mathrm{rad}$.

	Jac. diff.				Hermite diff.			
cutoff frequency	ω_0	$0.8\omega_0$	$0.7\omega_0$	$0.6\omega_0$	ω_0	$0.8\omega_0$	$0.7\omega_0$	$0.6\omega_0$
delay	6	7	8	9	9	12	14	17
window length	17	22	27	31	28	36	40	47
SNRout in dB	13.53	18.01	18.43	20.14	14.47	17.23	18.84	20.77
SBR in dB	44.08	39.16	35.12	31.64	34.84	34.8	31.62	26.21

at ω_c decreases with increasing β. It also approaches that of the Lag. diff. Increasing β also increases the estimation delay. For instance, for $\beta = 45$ the delay is eight sampling periods, which is the same as for the Lag. diff.

Decreasing the cutoff frequency of the Jac. diff. and keeping the filter order constant enhances the approximation results as shown in Fig. 3.14, where time evolution of the signals for different values of ω_c are given. From Table 3.3 it follows that when ω_c is decreased by a factor of 0.4 the SNRout increases from $13.53\,\mathrm{dB}$ to $20.14\,\mathrm{dB}$ which is comparable to that of the considered Lag. diff. Contrarily, the SBR decreases from $44.08\,\mathrm{dB}$ to $31.64\,\mathrm{dB}$. These results are not surprising since the filter window length T decreases with increasing ω_c as it can be seen from (3.6). Unsurprisingly, this confirms the theoretical results from Corollaries 2.4 and 2.5. For a lower cutoff frequency the delay also increases since it is proportional to T. For $\omega_c = 0.6\omega_0$ and a SNRout comparable to that of a Lag. diff. the delay is nine sampling periods compared to eight for the latter.

Decreasing the cutoff frequency of the Hermite diff. will increase the SNRout and decrease the SBR as given in Table 3.3. For a cutoff frequency $\omega_c = 0.6\omega_0$ the SNRout equal to $20.77\,\mathrm{dB}$ is comparable to that of the Jacobi and Laguerre diffs. with the cutoff frequencies $0.6\omega_0$ and ω_0, respectively. The SBR of these three diffs. are also comparable. However, the delay of the Hermite diff. equal to 17 sampling periods is twice as high as for the latter two. Thus, in order to achieve a comparable SNRout, the Hermite diff. yields a higher estimation delay.

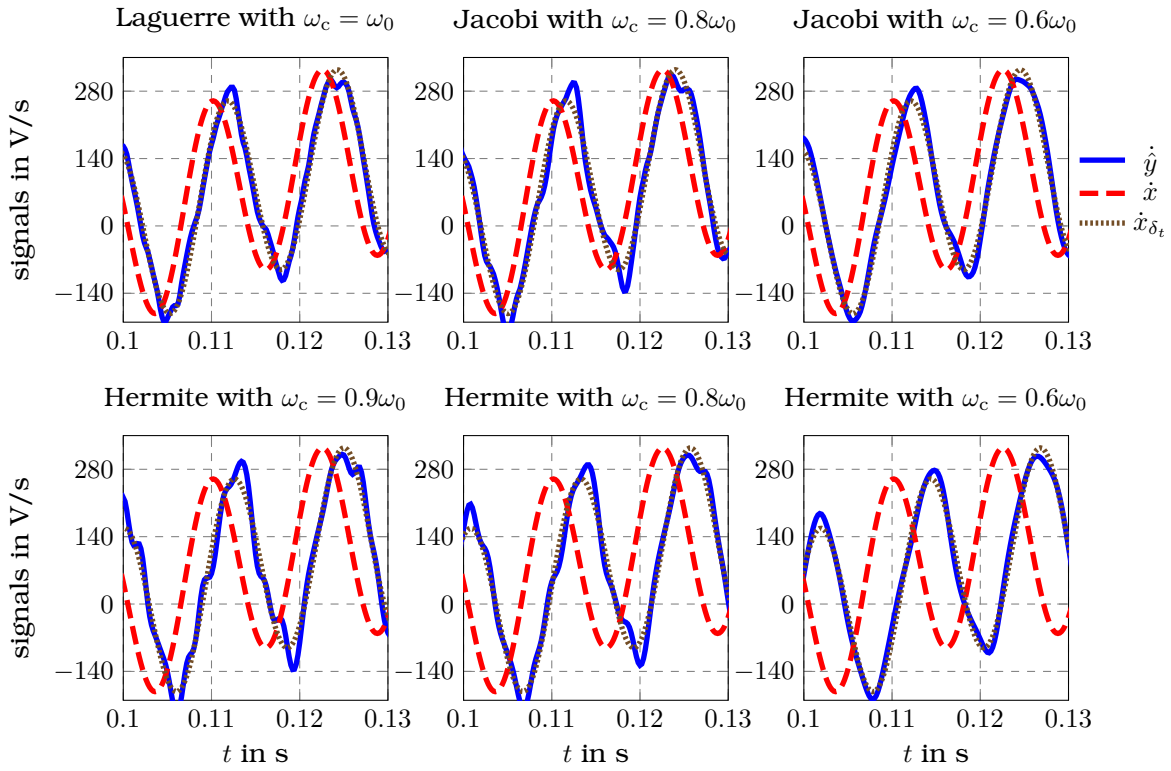


Figure 3.14: Approximation results for the first-order derivative of the signal x from Example 3.8 using a Laguerre, two Jacobi and three Hermite diffs. with different cutoff frequencies. The parameters of the diffs. are summarised in Tables 3.1 and 3.3.

It has already been observed in Section 2.2.2 that the DoE is reduced when delay-free approximations are considered. In Section 2.4 it has been observed that the amplitude spectra of delay-free diffs. exhibit significant overshoots which may result in bad approximation results. Fig. 3.15 illustrates the experimental results for the estimation of the first-order derivative of the signal y using delay-free Jacobi and Laguerre diffs. All parameters of the diffs. except ϑ are the same as those for the scenario in Table 3.1 and Fig. 3.12. As expected, a significant deterioration of the result can be observed. The SBR of the Jac. diff., for example, is equal to $7.09\,\mathrm{dB}$ as opposed to $44.08\,\mathrm{dB}$ when a delay is tolerated. Similarly, for the Lag. diff. the SBR is reduced from $24.82\,\mathrm{dB}$ to $2.57\,\mathrm{dB}$. The SNRout are reduced from $13.53\,\mathrm{dB}$ to $10.26\,\mathrm{dB}$ and from $20.0\,\mathrm{dB}$ to $19.5\,\mathrm{dB}$ for the Jac. diff. and the Lag. diff., respectively.

Comparison of computational and memory requirements Consider the approximation results given in Tables 3.1 and 3.4. For an identical cutoff frequency and the approximation of the diffs. as FIR filters, the Jac. diff. has the lowest memory and computational requirements. Only 17 filter coefficients and past measurements have to be saved as compared to 55 and 28 for the Laguerre and Hermite diffs., respectively. Thus, for the same cutoff frequency, the Lag. diff. requires to store approximately 323% more measurement values. This implementation, of course, immediately influences the computational requirements, which also increase by 323%. For the Hermite diff. these values increase by 196%. Thus, for the same cutoff frequency, the Jac. diff. is advantageous from a memory and computational perspective.

If the parameter β is increased to 45 such that the SNRout and SBR of the Jac. diff. are

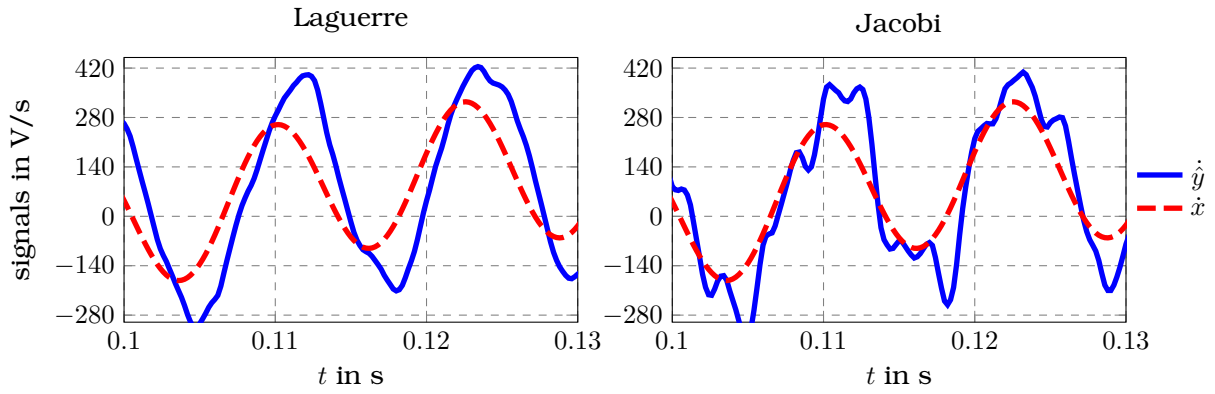


Figure 3.15: Delay-free approximation results for the first-order derivative of the signal x from Example 3.8.

Table 3.4: Computational and storage requirements for the approximation of the first-order derivative of the signal from Example 3.8 for $N = 1$ and $\omega_c = 8\pi f_2 \mathrm{rad}$. All diffs. are parametrised such that ϑ is a zero of the OP of interest with degree $N + 1$. The zero is chosen such that the delay is minimised. The abbreviations C and R stand for the discrete convolution and the recursive implementation, respectively. The X stands for no recursive implementation.

		Hermite diff.	Lag. diff.	Jac. diff.
		$\mu = 90\,\mathrm{ms}$, $T = 0.91\,\mathrm{ms}$	$\alpha = 3$, $T = 0.6\,\mathrm{ms}$	$\alpha = \beta = 3$, $T = 3.6\,\mathrm{ms}$
# filter coeff.	C.	56	110	35
	R.	X	11	21
# arith. opt.	C.	55	109	34
	R.	X	11	28
# past values y	C.	28	55	17
	R.	X	5	23
# past values $\hat{\dot{y}}$	C.	0	0	0
	R.	X	5	7

comparable to those of the Lag. diff. the computational and memory requirements of both diffs. implemented as FIR filters are comparable as it can seen by comparing the results in Tables 3.1 and 3.2. If $\alpha = \beta$ and the cutoff frequency of the Jac. diff. is reduced to $\omega_c = 0.6\omega$ such that SNRout is comparable to that of the Lag. diff., the computational and memory requirements of the first are 50% less than those of the second, which is a clear advantage.

The requirements for the recursive implementation of the Jacobi and the Laguerre diffs. discussed in Sections 3.2.2 and 3.2.3, respectively, are summarised in Table 3.1 for equal cutoff frequencies. For the Jac. diff. the bilinear transform has been used for the discretisation and the computations. For the Lag. diff. eleven filter coefficients have to be stored compared to 21 for the Jac. diff. which corresponds to a difference of 50%. The Lag. diff. requires the storage of five past values of the filter output and the input, respectively. Contrarily, the Jac. diff. requires 23 and seven, respectively. Thus, for the Jac. diff. significantly more values have to be stored. Similar observations can be made for the number of arithmetic operations. While

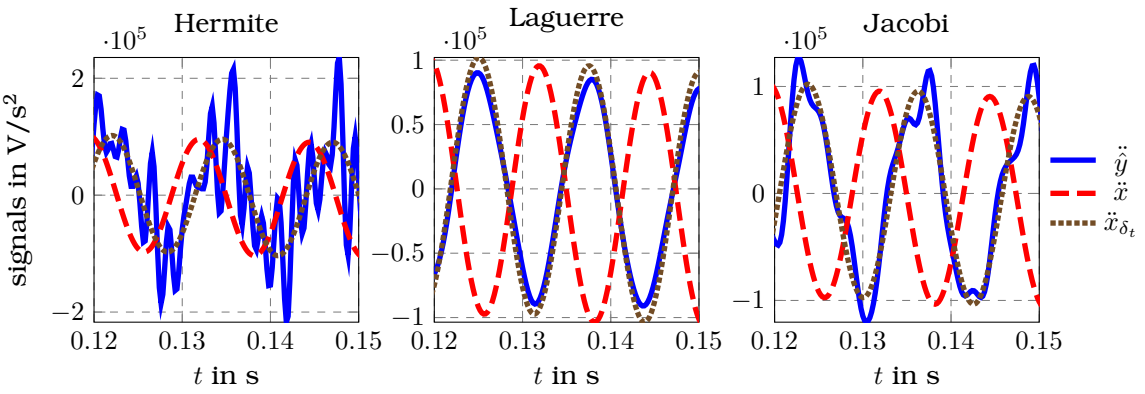


Figure 3.16: Approximation results for the second-order derivative of the signal x from Example 3.8. The parameters of the diffs. are summarised in Table 3.5.

for the Lag. diff. only eleven additions and multiplications are needed in total, the Jac. diff. requires 28, which corresponds to a factor of 2.54. Moreover, the recursive implementation of the Jac. diff. has to be performed in fixed-point arithmetic and requires a careful choice of the initial conditions as discussed in Section 3.2.3. The recursive implementation of the Lag. diff. does not suffer from these limitations.

When the Jac. diff. is implemented recursively with a cutoff frequency $\omega_c = 0.6\omega_0$ to achieve the same SNRout as for the Lag. diff. with $\omega_c = \omega_0$, the number of required arithmetic operations is the same as above. As discussed in Section 3.2.3, this does not depend on the length of the filter window. However, more filter coefficients, past measurements, and past filter values must be stored.

Approximation of the second-order derivative

Comparison of approximation results Start with diffs. having the cutoff frequency ω_0 and $\alpha = 12$ for the Lag. diff. and $\alpha = \beta = 12$ for the Jac. diff. The remaining parameters are summarised in Table 3.5. Fig. 3.16 shows the approximation results when all diffs. have the same cutoff frequency. The numerical values for the SNRout and the SBR are summarised in Table 3.5. As for the approximation of the first-order derivative, the Lag. diff. shows the highest SNRout and the lowest SBR. Thus, these diffs. show the best robustness with respect to the measurement disturbance but the largest error stemming from the truncation of the generalised Fourier series.

Decreasing the cutoff frequency of the Jacobi and the Hermite diffs. yields larger values for the SNRout as given in Table 3.6. The evolution of the signals is shown in Fig. 3.17, where the improvement of the approximation quality can be clearly seen. A decrease of the cutoff frequency from ω_0 to $0.4\omega_0$ for the Jac. diff., for example, results in a SNRout equal to $20.24\,\mathrm{dB}$, which is comparable to the $21.27\,\mathrm{dB}$ for the considered Lag. diff. However, the SBR of the first is larger than that of the second. For this choice of the cutoff frequency, the Jac. diff. has a delay of 37 sampling periods compared to the Lag. diff. with 28. This corresponds to a difference of approximately 25%. Parametrising the Jac. diff. such that the delay is equal to that of the Laguerre one yields a SNRout of $14\,\mathrm{dB}$ which is significantly worse than that of the Lag. diff.

Similar conclusions can be drawn for the Hermite diff. Decreasing the cutoff frequency

Table 3.5: Approximation results for the second-order derivative of the signal from Example 3.8 for $N = 1$, $\omega_c = 8\pi f_2 \text{rad}$. All diffs. are parametrised such that ϑ is the zero of the OP of interest with degree $N + 1$. The delay and the FIR window length are given as multiples of the sampling period t_s.

	Hermite diff.	Lag. diff.	Jac. diff.
	$\mu = 91\,\text{ms}$, $T = 0.91\,\text{ms}$	$\alpha = 12$, $T = 0.5\,\text{ms}$	$\alpha = \beta = 12$, $T = 10\,\text{ms}$
delay	13	28	21
window length T_L	34	97	52
SNRout in dB	−0.13	21.27	9.61
SBR in dB	39.01	16.43	28.61

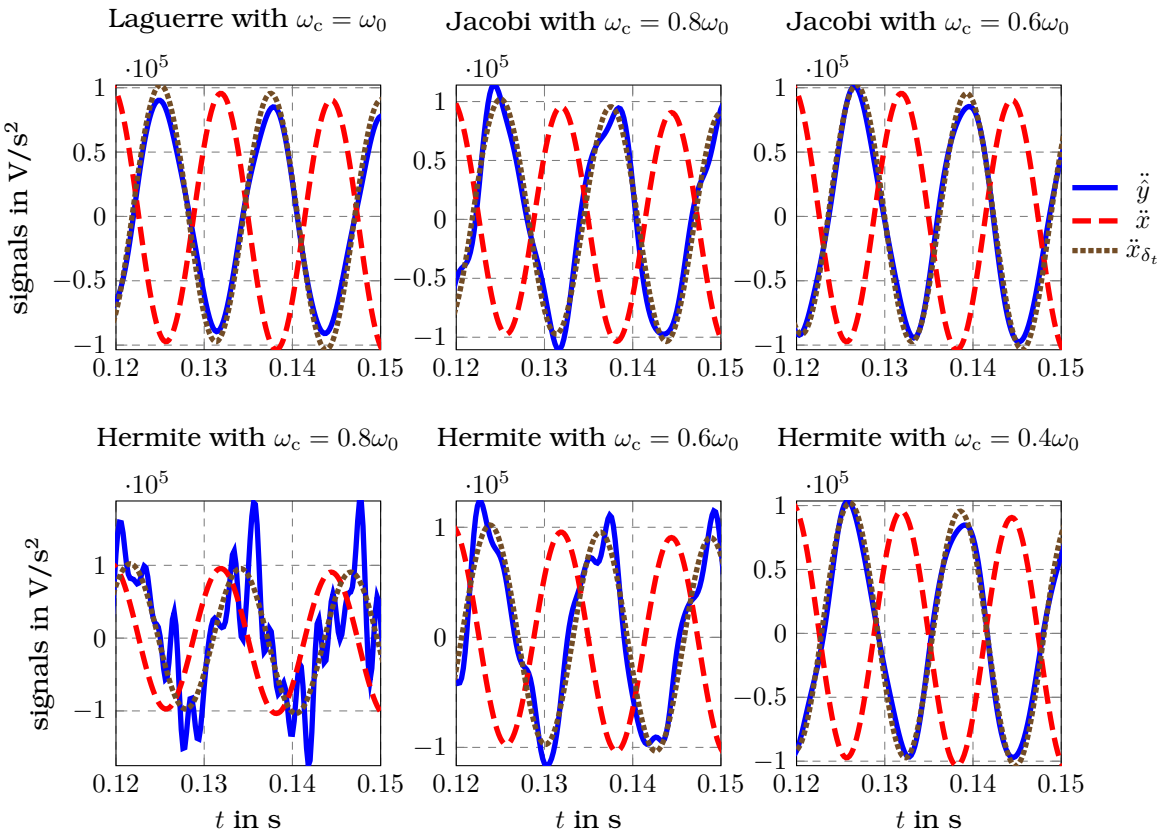


Figure 3.17: Approximation results for the second-order derivative of the signal x from Example 3.8 using a Laguerre, two Jacobi and three Hermite diffs. with different cutoff frequencies. The parameters of the diffs. are summarised in Tables 3.5 and 3.6.

significantly increases the quality of the approximation as shown in Table 3.6 and Fig. 3.17. To achieve a SNRout comparable to that of the Lag. diff. a higher delay must be taken into account.

Comparison of computational and memory requirements As for the approximation of the first-order derivative, the analysis of the computational and memory requirements are considered for the implementation as an FIR filter first. From Tables 3.5 and 3.6 it is clear

Table 3.6: Approximation results of the second-order derivative of the signal from Example 3.8 for Jacobi and Hermite diffs. with different cutoff frequencies and $N = 1$. For the Jac. diff. $\alpha = \beta = 12$. The delay and the FIR window length are given as multiples of the sampling period t_s and $\omega_0 = 4\pi f_2 \mathrm{rad}$.

	Jac. diff.				Hermite diff.			
cutoff frequency	ω_0	$0.8\omega_0$	$0.7\omega_0$	$0.6\omega_0$	ω_0	$0.8\omega_0$	$0.6\omega_0$	$0.4\omega_0$
delay	21	28	32	37	13	16	22	33
window length	52	65	75	87	34	42	55	83
SNRout in dB	9.61	14.00	16.99	20.24	-0.13	4.7	10.40	18.75
SBR in dB	28.61	26.60	26.14	21.32	39.01	33.77	24.71	22.14

that in order to achieve comparable values for the SNRout all three diffs. have similar filter window lengths. For instance, for the Jac. diff. it is 87 sampling periods when the cutoff frequency is $0.6\omega_0$ as compared to the 97 of the Laguerre one. for the Hermite diff. it is 83. Thus, the computational and the memory requirements are comparable.

However, when the Lag. diff. is implemented as an IIR filter using the bilinear transform, it is straightforward to conclude that the numbers of required arithmetic operations and measurement values to be stored are drastically reduced. For instance, only 14 past values of the measurement and of the approximated derivative need to be stored in comparison to the 97 for the FIR implementation. This corresponds to a reduction of 85% of the memory requirements. Only 57 arithmetic operations need to be performed at each instant as opposed to the 194 for the discrete convolution, which corresponds to a reduction of 70%. Similarly, for the Jac. diff. a recursive implementation reduces the number of required operations and measurement values to be stored. This approach, however, requires to be implemented in fixed-point arithmetic and a proper initialisation is necessary.

3.3.2 Approximate inversion of analogue anti-aliasing filters

Often in practical applications[4], the achievable sampling rates of the real-time hardware are limited for technical or economical reasons and, thus, analogue low-pass filters are required in order to avoid irreparable signal distortions due to aliasing (see, e.g., Kotelnikov (1933); Nyquist (1928); Shannon (1948, 1949)). Unfortunately, these anti-aliasing filters also restrict the dynamics of the sampled signals available for embedded control, state estimation, or diagnosis algorithms, which can deteriorate the achievable system performance.

One way to overcome this is to improve the electronic hardware, e.g, by using a faster computation unit which allows to increase the sampling frequency or by using a more complicated filter circuit with a higher stopband slope. Such measures increase the product costs. As an alternative, a bank of algebraic numerical diffs. can be used for an approximate inversion of the analogue anti-aliasing filter, as proposed in Kiltz et al. (2021, 2019). By doing so, an improved signal dynamics and a simultaneous estimation of signal derivatives can be

[4]The theoretical part of this subsection has first been published in Othmane et al. (2022). The experimental validation and the comparison of the different diffs. are first published here.

achieved without the need for a higher sampling rate or more complicated filter circuits. This is demonstrated in the following.

The input-output characteristics of typical analogue anti-aliasing filters (from simple RC circuits to higher-order Butterworth, Bessel, Chebyshev or elliptical filters) can be modelled as a linear time-invariant differential equation

$$\sum_{i=0}^{n} a_i y^{(i)}(t) = u(t)$$

with the input signal $u : t \mapsto u(t)$, the output signal $y : t \mapsto y(t)$, and the known scalar filter coefficients a_i, $i = 0, \ldots, n$, where $n \in \mathbb{N}$ is the filter order (see, e.g., (Tietze, Schenk, & Gamm, 2008, Ch. 13)).

In a practical application, the filter output y is sampled by the real-time hardware, whereas the filter input u is unknown. In contrast to this, it would often be desirable for the real-time algorithms to rely on the filter input u (or on one or more time derivatives $u^{(m)}, m \in \mathbb{N}$, of u) instead of the filter output y. In such a case, an approximation $\hat{u}^{(m)}$ of the sought signal $u^{(m)}$ can be generated from the measurement y by

$$\hat{u}^{(m)}(t) = \sum_{i=0}^{n} a_i \left(g^{(i+m)} \star y \right)(t),$$

with a diff. g appropriately designed and $g^{(i+m)} \star y$ the convolution of $g^{(i+m)}$ with y.

Note that the filter characteristics of the diff. can be chosen independently of the filter characteristics of the analogue anti-aliasing filter. In particular, the cutoff frequency of the diff. can be chosen higher and its stopband slope steeper than the respective characteristics of the anti-aliasing filter, which explains the achievable increase in the signal dynamics. This method has successfully been applied in Kiltz et al. (2019) for the reconstruction of the first-order and second-order derivatives of a measured current signal in an automotive series product with lean embedded hardware.

Practical limitations of the realisable signal dynamics are due to the fact that the discrete-time approximations of the convolution integrals $g^{(i+m)} \star y$, $i = 0, \ldots, n$, must be sufficiently accurate. See Section 3.2 for suitable discretisation methods and their limitations depending on the sampling rate.

A drawback of the proposed method is that, for the sake of sufficient noise suppression, a delay in the reconstructed input signal or its derivatives must be accepted. This delay tends to increase with the required order m of the reconstructed derivative, the filter order n of the anti-aliasing filter, and the cutoff frequency and stopband slope of the diff. used, i.e., the desired bandwidth of the reconstructed signal. Nevertheless, given the parametrisation rules in Section 3.1, often an acceptable compromise between the signal delay and the bandwidths of the reconstructed signals can be found. In particular, the signal delay might be of minor interest in comparison with the increased resolution of the reconstructed signals especially for applications such as fault diagnosis, anomaly detection or event detection, as in the case study in Kiltz et al. (2019).

Approximate inversion of a first order filter using experimental measurements

In the following, the method is applied to reconstruct the input signal of a single-stage anti-aliasing filter

$$\tau \dot{y}(t) + y(t) = u(t)$$

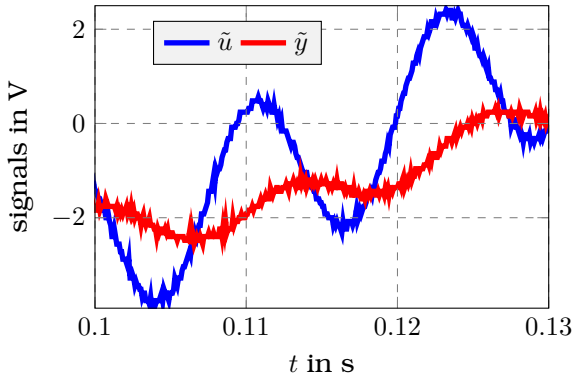


Figure 3.18: Evolution of the measured input and output of an analogue RC circuit with cutoff frequency 19.40 Hz.

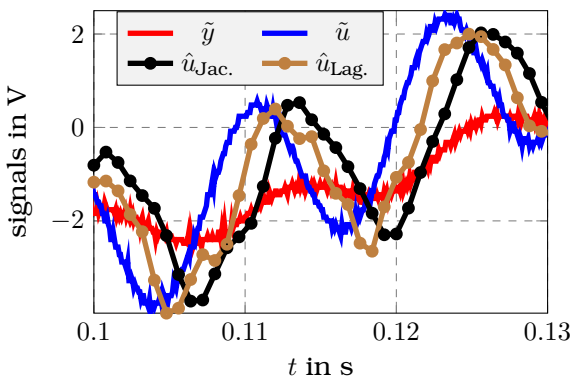


Figure 3.19: Evolution of the measured input $\tilde{u}$ and its approximations using the considered Jacobi and Laguerre diffs.

realised by an RC circuit. The time constant of the anti-aliasing filter is $\tau = 8.2\,\text{ms}$ (which corresponds to a $-3\,\text{dB}$ cutoff frequency of $19.40\,\text{Hz}$), and the sampling period for the measurement of y is $t_\text{s} = 0.8\,\text{ms} \approx \tau/10$. The filter is driven by the input

$$u(t) = a_1 \sin(2\pi f_1 t) + a_2 \sin(2\pi f_2 t), \quad f_1 = 10\,\text{Hz},\ f_2 = 80\,\text{Hz},\ a_1 = 2.5\,\text{V},\ a_2 = 1.75\,\text{V}.$$

To compare the approximation results, the input u is measured with a sampling period of $0.01\,\text{ms}$. The evolution of the disturbed measurements $\tilde{u}$ and $\tilde{y}$ of u and y, respectively, are given in Fig. 3.18.

The signal u is reconstructed using Laguerre and Jacobi diffs. with delays. The parameters $N = 1$ and $\alpha = \beta = 5$ are chosen for both diffs. Fig. 3.20 shows the amplitude and phase spectra of a Laguerre and a Jacobi diffs. with cutoff frequencies equal to $2356.19\,\text{rad/s}$ and $1220.09\,\text{rad/s}$, respectively. These values are chosen such that both diffs. have comparable properties for low frequencies. The discrete filters preserve the frequency-domain properties and no aliasing phenomena can occur with these values. The error norm in (3.12) is below $-60\,\text{dB}$ for the highest derivative order, for both diffs.

From Fig. 3.20 it can be seen that for low frequencies the phase spectrum of both diffs. show a linear dependency on the angular frequency and the phase Jac. diff. is lower than that of the Lag. diff. Thus, a higher delay has to be expected from the former. This is confirmed by the results in Table 3.7: While the Lag. diff. has a delay of two sampling periods, that of the Jac. diff. is equal to three. The reconstructed signals are shown in Fig. 3.19 and confirm this analysis. Both approaches yield convincing results and a very effective reconstruction of the original signal and a good noise suppression can be achieved at the cost of comparably small signal delays.

The Jac. diff. has a window length of eleven sampling periods that is significantly smaller than the Lag. diff. with 19. However, an IIR approximation of the Lag. diff. can significantly decrease the number of required arithmetic operations and the measurement values to be stored as discussed in the previous case studies from Section 3.3.

3.4 Automatic tuning of orthogonal differentiators

When designing a diff., a significant issue is the lack of a possibility to compare the estimated signal with the actual derivative, which is unknown. The systematic tuning approaches

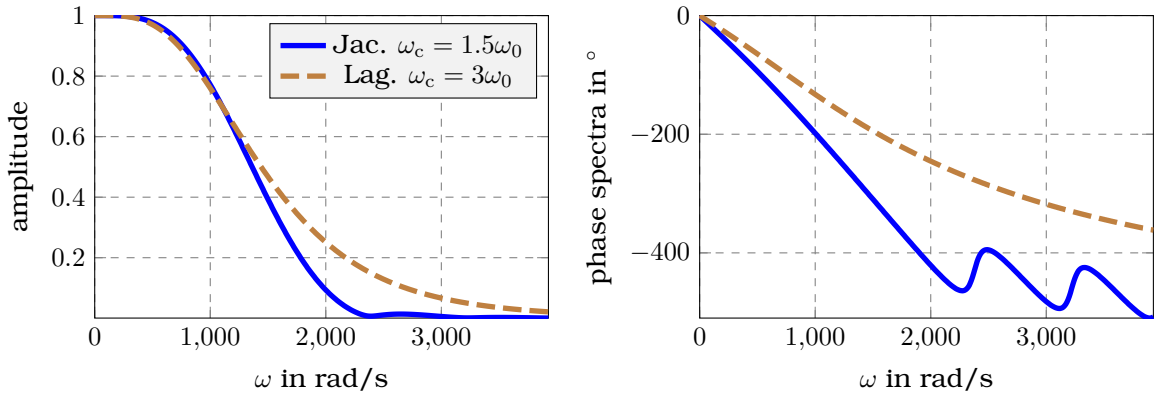


Figure 3.20: Amplitude and phase spectra of the diffs. used for the approximate numerical inversion of an analogue RC circuit. The parameter ω_0 is equal to $\frac{\pi}{10t_s}$.

Table 3.7: Results for the approximate inversion of a first-order analogue anti-aliasing filter using Jacobi and Laguerre diffs. The latter are approximated by FIR filters. The delay and the filter window lengths are given in multiples of the sampling period.

	window length	delay	cutoff frequency
Lag. diff.	19	2	2356.19 rad/s
Jac. diff.	11	3	1220.09 rad/s

discussed in Section 3.1 have been developed separately for each diff. by considering the low-pass properties of the filters. A more general and automatic tuning approach requiring only the design of one parameter is proposed here. It relies on the filter interpretation of the diffs.

3.4.1 Proposed approach

Let $g_{N,\vartheta}$ be an orthogonal differentiator (orth. diff.) of order n as defined in Definition 2.5. Denote by $\hat{f}$ and $\hat{f}^{(n)}$ the approximations of f and $f^{(n)}$, respectively, computed using the same diff. $g_{N,\vartheta}$. The offline tuning is considered first and the generalisation to an online adaptive tuning is considered later. Assume that f is known in the interval $[0, t^*]$, $t^* > 0$, square integrable, and

$$\mathcal{P}(f) = \frac{1}{t^*} \int_0^{t^*} |f(\tau)|^2 \, \mathrm{d}\tau \neq 0.$$

Recall from Table 2.4 that the approximated derivative can be seen as the output of the linear time-invariant filter with impulse response $g_{N,\vartheta}$ having as input the sought n-th order derivative $f^{(n)}$ of f. In numerous applications the filtered version of the function f itself is also required. Thus, f and its approximation $\hat{f}$ shall be used to design the diffs. The new parametrisation approach relies on the comparison of the filter output, i.e., the approximated function, with the signal itself. Since f and the sought derivative are filtered by the same filter, the approximation of f is used to design a suitable cost function.

The filter shall now be designed such that the ratio

$$\kappa_{\mathcal{P}} = \frac{\mathcal{P}(\hat{f})}{\mathcal{P}(f)} \tag{3.29}$$

is equal to a desired value $\kappa_{\mathcal{P}}^* \in (0, 1]$. Thus, the filter shall attenuate the power of the input by the factor $\kappa_{\mathcal{P}}^*$. Denote in the following by p the tuple of non-integer parameters on which the diff. $g_{N,\vartheta}$ depends. For the Jacobi and Laguerre diffs. the corresponding tuples are $p_{\text{Jac.}} = (\alpha, \beta, T)$ and $p_{\text{Lag.}} = (\alpha, T)$. For the Hermite diff. only the parameter T is considered. Thus, $p_{\text{Her.}} = T$. By choosing μ large enough the FIR approximation discussed in Section 3.2.1 yields the filter window length and the corresponding delay and μ does not need to be optimised.

By keeping in mind the effects of discretisation, the diff. can be designed by solving the following optimisation problem:

$$\min_{p \in \mathbb{P}_n} \quad \frac{|\kappa_{\mathcal{P}}(p) - \kappa_{\mathcal{P}}^*|}{\kappa_{\mathcal{P}}^*} \tag{3.30a}$$

$$\text{subject to} \quad \mathcal{J}(n, p) \leqslant \mathcal{J}_{\max} \tag{3.30b}$$

$$\frac{\omega_{\text{N}}^n \left|\mathcal{F}\left\{g_{N,\vartheta}\right\}(\omega_{\text{N}})\right|}{\omega_{\text{c}}^n \left|\mathcal{F}\left\{g_{N,\vartheta}\right\}(\omega_{\text{c}})\right|} \leqslant k_{\text{N},\max}. \tag{3.30c}$$

In the latter optimisation problem, $\mathbb{P}_n$ and $\kappa_{\mathcal{P}}(p)$ are the set of admissible parameters for the estimation of the n-th order derivative and the ratio from (3.29) computed using a diff. with parameters p, respectively. For a Lag. diff. with a delayed approximation and a fixed polynomial degree N, the parameters are $p_{\text{Lag.}} = (\alpha, T)$ and[5] $\mathbb{P}_n = \mathbb{R}_{>n-1} \times \mathbb{R}_{>0}$, for example. The quantity $\mathcal{J}(n, p)$ in (3.30b) denotes the error norm defined in (3.12) computed for a diff. approximating the n-th order derivative using the parameters p. Choosing $\mathcal{J}_{\max}$ small ensures that the discrete differentiator preserves the frequency-domain properties of the continuous-time one. In the second constraint (3.30c), ω_{N}, $\mathcal{F}\left\{g_{N,\vartheta}\right\}$ and $k_{\text{N},\max}$ denote the Nyquist frequency, the Fourier transform of the differentiator and the weakest relative attenuation the differentiator should have at the Nyquist frequency, respectively. Choosing $k_{\text{N},\max}$ small ensures that aliasing problems are avoided. Thus, the approach finds parameters that minimise the relative deviation of the ratio in (3.29) from a desired value $\kappa_{\mathcal{P}}^*$ such that the discretisation does not alter the approximation results.

The approach is now first applied for the approximation of the first-order and second-order derivatives of a measured signal from Example 3.8. The experimental validation of this design approach shows promising results as will be shown in the next section. The theoretical aspects are then discussed in the concluding remarks of this section.

Remark 3.1: Considering a maximally tolerable delay

The optimisation (3.30) can be further extended to include a maximal value for the estimation delay. This can be important when the approximated derivatives are used in closed-loop control.

[5] The set $\mathbb{R}_{>x}$ is defined as $\mathbb{R}_{>x} = \{y \in \mathbb{R} | y > x\}$.

Remark 3.2: Motivation for the approach

This design approach is motivatived by the concept of power gain used for the design of electronic circuits (see, e.g., (Pettit & McWhorter, 1961, Ch. 1) and (Egan, 2003, App. G)). For the design of amplifiers where the purpose of the circuit is the amplification of the level of the input signal, for example, the amount of increase is known as gain of the amplifier.

3.4.2 Examples for automatic tuning

The proposed approach is now applied to approximate the first and second-order derivatives of the signal from Example 3.8. Recall that this signal has already been used to discuss the systematic tuning guidelines developed in Section 3.1. For all diffs. the FIR implementation is used, $\mathcal{J}_{\max} = -60\,\mathrm{dB}$ and $k_{\mathrm{N,min}} = -50\,\mathrm{dB}$. Thus, the discrete diff. preserves well the frequency domain properties of the continuous one. The windows of the filters are computed such that the window starts when the distribution function takes values greater than 10^{-4}. For the optimisation the entire history of the signal from Example 3.8 is used. All optimisation problems are solved using the interior-point method (see Byrd, Hribar, and Nocedal (1999)). The cost function at the optimal solution is always below 10^{-5}.

Estimation of a first-order derivative

Fig. 3.21 shows the variation of $\kappa_{\mathcal{P}}$ from (3.29), the estimation delay, and the length of the filter window as a function of the logarithm of the parameter T of the Hermite diff. The values that follow from the solution of the optimisation problem are denoted by a star in the superscript and shown in the latter figure. Increasing T decreases $\kappa_{\mathcal{P}}$ and increases the delay and the window length. These observations also follow from the theoretical considerations in Section 3.1. Thus, a global solution for the optimisation problem exists.

Fig. 3.22 shows the variation of $\kappa_{\mathcal{P}}$, the estimation delay, and the length of the filter window as a function of the parameters T and α of the Lag. diff. It can be seen that level sets exists, where $\kappa_{\mathcal{P}}$ is constant for different values of T and α. Thus, no global solution can be found using the interior-point method. The variation of the quantities for the Jac. diff. is not depicted. However, here again, numerous local minima have been, and no global solution exists.

Table 3.8 gives the optimal diff. parameters, when the optimisation process is initialised with parameters subject to a difference of 30% from those determined using the systematic tuning guidelines in Section 3.3.1. The desired value of the attenuation is $\kappa^*_{\mathcal{P}} = 0.99$. The table also summarises the estimation delay, the cutoff frequency, the filter window length, the SNRout, and the SBR. For the Hermite diff., the cutoff frequency is computed as the frequency where the amplitude reaches $-3\,\mathrm{dB}$. Fig. 3.24 shows the resulting amplitude spectra that are very close to each other for frequencies up to $1000\,\mathrm{rad/s}$. A better disturbance attenuation has to be expected from the Jac. diff. due to the local minima of the spectrum at approximately $1750\,\mathrm{rad/s}$, for example, for the first one. The worst disturbance attenuation should follow for the Hermite diff. since its amplitude spectrum is up to $2000\,\mathrm{rad/s}$ higher than that of the other two.

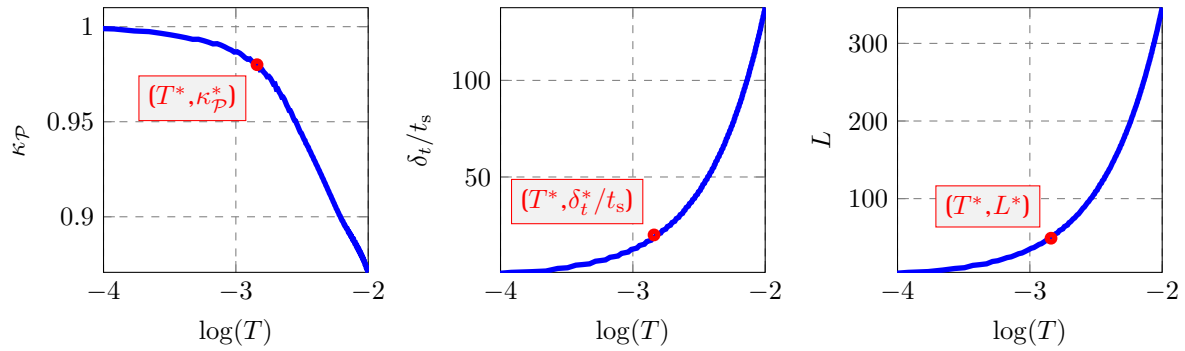


Figure 3.21: Variation of the ratio $\kappa_{\mathcal{P}}$ defined in (3.29), the estimation delay and the number of filter coefficients L as a function of the parameter T of a Hermite diff. for the estimation of the first-order derivative from Example 3.8. The values resulting from the optimal parameter T^* are given by the red dots and summarised in Table 3.8.

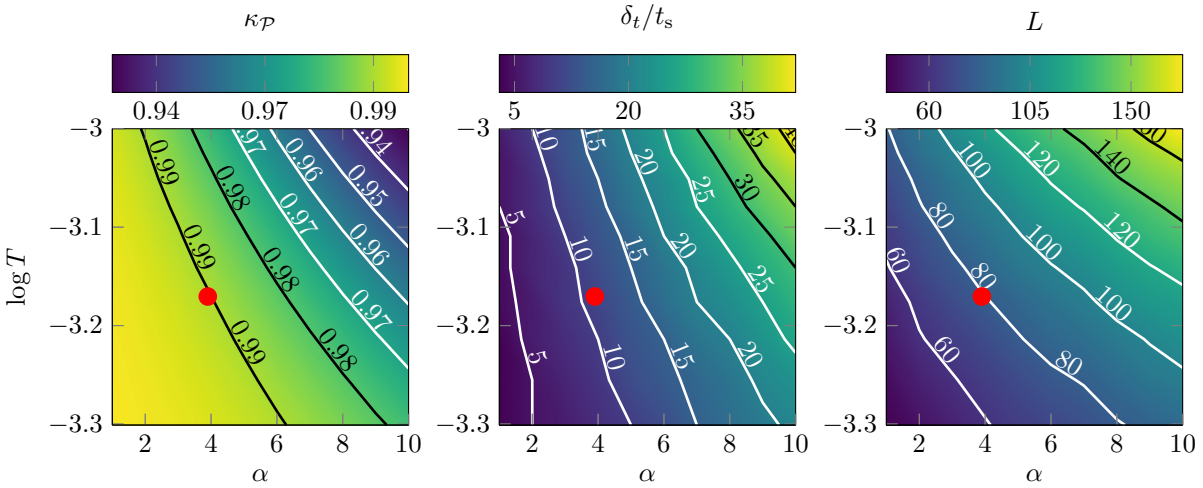


Figure 3.22: Variation of the ratio $\kappa_{\mathcal{P}}$ defined in (3.29), the estimation delay and the number of filter coefficients L as a function of the parameter T and α of a Lag. diff. for the estimation of the first-order derivative from Example 3.8. The values resulting from the optimal parameters are given by the red dots and summarised in Table 3.8.

From Table 3.8, the Laguerre and the Jacobi diffs. have the same estimation delay of 11 sampling periods. That of the Hermite diff. is 18. The Jac. diff. has the highest SNRout equal to 25.05 dB and the Hermite diff. the lowest equal to 20.68 dB. This confirms the previous analysis of the amplitude spectra. However, the time evolution of the approximations given in Fig. 3.23 show very good results for all cases. Comparing these results with those obtained using the systematic tuning summarised in Tables 3.1 and 3.3 shows that by accepting a slightly higher delay the quality of the approximations increases significantly. For the Jac. diff. the delay for this parametrisation is only two sampling periods higher than that in Section 3.3.1 for $\omega_c = 0.6\omega_0$.

While the Jacobi and Hermite diffs. have a window length of $35t_s$ and $34t_s$, respectively, the Laguerre one requires $67t_s$. Thus, computing the approximation with the latter requires storing more filter coefficients and measurement values. However, approximating the optimal Lag. diff. by a Lag. diff. that has the same cutoff frequency but the order $\bar{\alpha} = 4$, which corresponds to the closest integer to the optimal α, yields an SNRout of 22.95 dB, the same

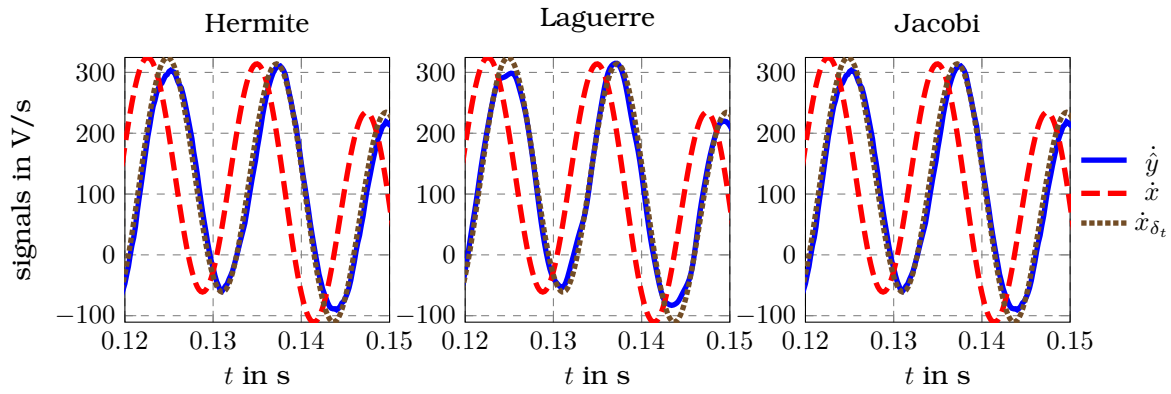


Figure 3.23: Approximation results for the first-order derivative of the signal x from Example 3.8 using diffs. whose parameter are computed using the automatic tuning approach. The parameters of the diffs. are summarised in Table 3.8.

Table 3.8: Approximation results for the first-order derivative of the signal in Example 3.8 for $N = 1$, $\omega_c = 8\pi f_1 \mathrm{rad/s}$ using the automatic tuning approach with $\kappa^*_{\mathcal{P}} = 0.99$. All diffs. are parametrised such that ϑ is the zero of the $N + 1$ OP of interest. The delay and the filter window length are given as multiples of the sampling period t_s.

	Hermite diff. $\mu = 1.6\,\mathrm{ms}$, $T = 1.5\,\mathrm{ms}$	Lag. diff. $\alpha = 3.898$, $T = 0.66\,\mathrm{ms}$	Jac. diff. $\alpha = 1.467, \beta = 1.216$ $T = 7.25\,\mathrm{ms}$,
cutoff freq. in rad/s	1215.18	1816.02	476.23
delay δ_t	18	11	11
window length T_L	34	67	35
SNRout in dB	20.68	22.82	25.05
SBR in dB	26.04	24.86	23.76

delay, and equal SBR. The cost (3.30) for this diff. is equal to $6 \cdot 10^{-4}$, which is close to that of the optimal solution. Thus, using the IIR implementation of this diff. significantly reduces the computational burden and the storage requirements without altering the results.

Estimation of a second-order derivative

Consider again the estimation of the second-order derivative of the signal from Example 3.8. Table 3.9 summarises the optimal parameters of the diffs. and the results of the derivative approximation process when $\kappa^*_{\mathcal{P}} = 0.99$. The time evolution of the approximations are shown in Fig. 3.25. The Jacobi and Laguerre diffs. have very close SNRout which are equal to $12.96\,\mathrm{dB}$ and $12.29\,\mathrm{dB}$, respectively, compared to that of the Hermite diff. having $10.17\,\mathrm{dB}$. The delay of Lag. diff. is equal to 15 sampling periods and is lower than those of the Hermite and Jacobi diffs. and equal to 18 and 20, respectively.

As before, the Lag. diff. has the largest filter window with length equal $69t_s$ as compared to the Jacobi and Hermite with $56t_s$ and $49t_s$, respectively. However, approximating the Lag. diff. by a Lag. diff. with the same cutoff frequency and a parameter α equal to the closest integer

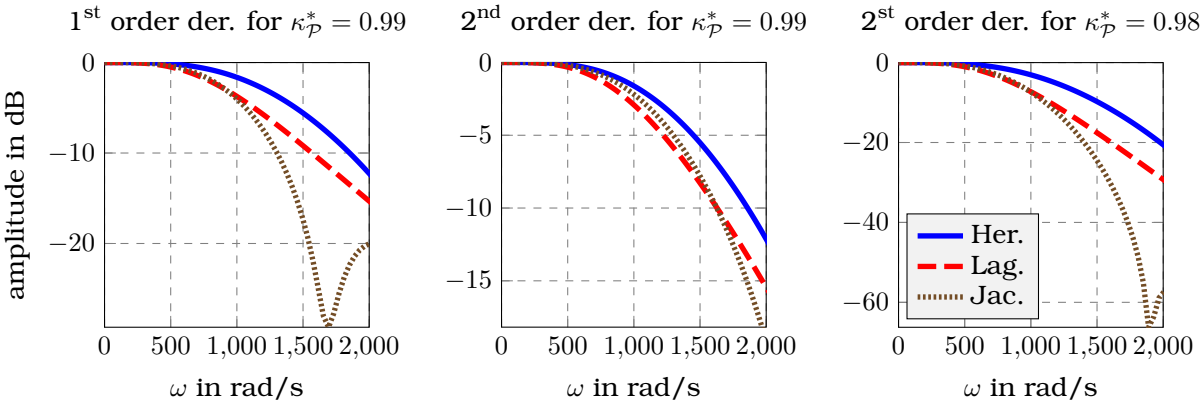


Figure 3.24: Amplitude spectra of the differentiators parametrised using the automatic tuning approach for the estimation of the first and second-order derivatives of the signal x from Example 3.8. The parameters of the diffs. are summarised in Tables 3.9 and 3.10.

Table 3.9: Approximation results for the second-order derivative of the signal in Example 3.8 for $N = 1$ using the automatic tuning approach with $\kappa_{\mathcal{P}}^* = 0.99$. All diffs. are parametrised such that ϑ is the zero of the $N + 1$ OP of interest. The delay and the filter window length are given as multiples of the sampling period t_s.

	Hermite diff.	Lag. diff.	Jac. diff.
	$\mu = 1.6\,\mathrm{ms}$, $T = 1.5\,\mathrm{ms}$	$\alpha = 8.93$, $T = 0.41\,\mathrm{ms}$	$\alpha = \beta = 6.04$ $T = 9.34\,\mathrm{ms}$
cutoff freq. in rad/s	1220.2	2728.91	1322.14
delay δ_t	20	15	18
window length T_L	49	69	56
SNRout in dB	10.17	12.29	12.96
SBR in dB	26.79	24.48	32.90

to the optimal solution, yields a delay of $15t_\mathrm{s}$ and a SNRout equal to $12.36\,\mathrm{dB}$ and the same SBR. The cost function for this diffs. is $1 \cdot 10^{-5}$, which is close to that for the optimal solution. Thus, a very efficient IIR implementation can be used without altering the results.

The approximation results when $\kappa_{\mathcal{P}}^* = 0.98$ are summarised in Table 3.10. It can be seen that the cutoff frequencies decrease in comparison with those computed for $\kappa_{\mathcal{P}}^* = 0.98$. This is not surprising since reducing the cutoff frequencies is required to attenuate the signals more. The time-evolution of the approximated derivatives is given in Fig. 3.25. Similar observations can be made regarding the delay, the SNRout and the quality of the approximation as before. A detailed discussion is thus omitted. Comparing the results with those achieved with the systematic tuning guidelines shows, for example, for the Jac. diff. that a higher SNRout can be achieved with a lower delay.

3.4.3 Concluding remarks

The previous examples show that the proposed approach for the automatic tuning of the differentiators is promising and reduces the number of free parameters to a single one. The

Table 3.10: Approximation results for the second-order derivative of the signal in Example 3.8 for $N = 1$ using the automatic tuning approach for $\kappa_{\mathcal{P}}^* = 0.98$. All diffs. are parametrised such that ϑ is the zero of the $N + 1$ OP of interest. The delay and the filter window length are given as multiples of the sampling period t_s.

	Hermite diff.	Lag. diff.	Jac. diff.
	$\mu = 1.9\,\mathrm{ms}$, $T = 1.84\,\mathrm{ms}$	$\alpha = 9.14$, $T = 0.58\,\mathrm{ms}$	$\alpha = 10.05, \beta = 9.92$ $T = 7.6\,\mathrm{ms}$,
cutoff freq. in rad/s	993.15	1939.57	1046.61
delay δ_t	25	22	34
window length T_L	59	97	94
SNRout in dB	14.33	19.31	21.24
SBR in dB	21.76	18.72	17.78

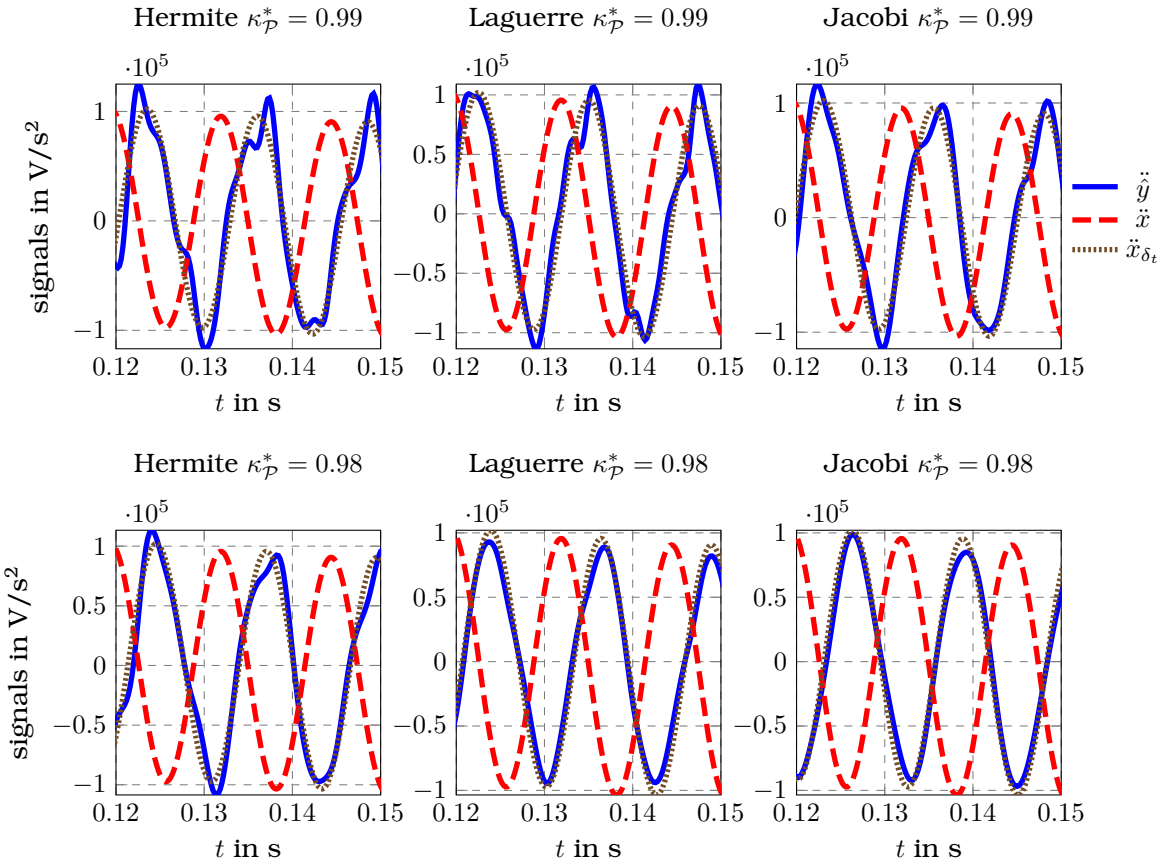


Figure 3.25: Approximation results for the second-order derivative of the signal x from Example 3.8 using diffs. with parameters computed using the automatic tuning approach for two values of $\kappa_{\mathcal{P}}^*$. The parameters of the diffs. are summarised in Table 3.9.

initialisation of the optimisation algorithms is, however, important due to the existence of multiple local minima. Thus, combining it with the systematic tuning guidelines yields excellent and promising results. Nevertheless, several questions are open and should be addresses in future research projects:

1. How to choose the parameter $\kappa_{\mathcal{P}}^{*}$?
2. How is the ratio $\mathcal{P}\left(\hat{f}^{(n)}\right)/\mathcal{P}\left(f^{(n)}\right)$ affected by this approach?
3. Which global optimisation scheme is the best adapted for solving the problem (3.30)?
4. How can this approach be used to design adaptive diffs.?

The last question is an interesting application of this design technique. The known history of a signal can then be used to adapt the parameters online.

3.5 Summary and concluding remarks

The tuning and discrete-time implementation of Hermite, Laguerre, and Jacobi diffs. have been discussed in detail in this chapter. A distribution function Φ defined in (3.9) has been used to derive an FIR implementation for the diffs. Systematic tuning guidelines have been derived such that desired low-pass properties can be achieved for the diffs. The effect of the discrete-time implementation on these desired properties has been analysed using an error norm defined in (3.12). Different discretisation schemes have been discussed. These properties and ideas have been illustrated using simple examples. For the Jacobi and Laguerre diffs., recursive implementations have been discussed. While those for the Jac. diffs. have already been published in Kiltz (2017), the discussions for the Laguerre ones are new.

For the Lag. diffs., the relation to HG diffs. and state variable filters has been discussed. It has been shown that the latter two established approaches are recovered for a particular parametrisation of the Lag. diffs. The simultaneous estimation of several derivatives using a filter having the same properties has been considered. Implementing these filters significantly reduces the required computational burden and the number of parameters and measurement points to be buffered.

The systematic tuning guidelines and the use of the diffs. have been illustrated for the approximation of derivatives of a measured signal and the approximate inversion of an analogue anti-aliasing filter. Using experimental data, it has been shown that compelling results can be achieved.

Finally, an automatic tuning approach based on an optimisation problem has been proposed. While the approach shows very convincing results on experimental data for all considered diffs., some theoretical questions remain open for future research.

The use of these diffs. for more elaborated problems is considered in the next chapters. It is shown that the approaches can be successfully used when higher-order derivatives are required.

4

Model-based detection of impulsive forces

The detection of abrupt changes in measured signals is crucial in numerous applications. For instance, popular models for sensor and actuator faults are stepwise changes (see, e.g., (Isermann, 2006, Ch. 5) and (Isermann, 2011, Ch. 2)). The ability to detect these abrupt changes is important to designing fault-tolerant control techniques. The use of numerical differentiation and in particular Jacobi differentiators (Jac. diffs.) for the detection of abrupt changes has been successfully demonstrated in simulations and experimental validations (see, e.g., Belkoura, Floquet, Ibn Taarit, Perruquetti, and Orlov (2011); Bourgeot and Delaleau (2007); Fliess, Join, and Mboup (2010); Kiltz (2017); Kiltz et al. (2014, 2012); Kiltz and Rudolph (2013); Tiganj and Mboup (2009)). For an extensive list of fault-tolerant control applications based on Jac. diffs., the reader is referred to Othmane et al. (2022). In this chapter, Jacobi and Laguerre differentiators (diffs.) are used to detect impulsive forces in two different experimental case studies.

First, the detection of collisions of a table tennis ball with a magnetically levitated plate and the estimation of the collision time is analysed in Section 4.1. The setup of the system can be found in Fig. 4.1. Only the measurement of the vertical displacement of the plate and the applied magnetic forces are known. The position of the ball is not measured. In addition to the corruption stemming from noise, the vertical displacement measurement is disturbed by a signal stemming from an audible mechanical eigenmode of the plate. This problem has been solved in Kiltz (2017); Kiltz et al. (2013); Kiltz and Rudolph (2013) using Jac. diffs. Therein, the parametrisation of the diffs. allowing the annihilation of undesired harmonic signals recalled in Section 3.1.3 has been used. The current work shows that using Laguerre differentiators (Lag. diffs.) in combination with a notch filter (see, e.g., (Tietze et al., 2008, Sec. 13.8) for a discussion of notch filters) yields better results. For instance, the approach allows to detect a ball falling from a lower height and considerably reduces the computational burden and storage requirements.

Second, Jac. diffs. are used in Section 4.2 for the data and computation efficient detection of faults for rolling element bearings. Numerical differentiation is combined with a physical model of the considered system to derive the required algorithms. Fault-free data can be used to identify the numerical values of the system parameters. Signals sensitive only to faults are designed using a simplified model of the bearing vibrations to detect the faults. These signals, called residuals, are proportional to the force stemming from the faults and are easy to compute and implement. The design of a failure detection threshold is discussed. When the residual value is above the threshold, a defect is detected. The desired probability of

false alarms can be specified when the threshold is designed. The method is validated using experimental data sets obtained from different research centres. The section has first been published in Othmane and Rudolph (2021)[1] and is taken literally.

Chapter content

4.1 Model-based collision detection on a magnetically supported plate 95
4.1.1 Collision detection and collision time estimation 96
4.1.2 Parametrisation of differentiators 98
4.1.3 Implementation of detection algorithms 101
4.1.4 Concluding remarks 103
4.2 Model-based fault detection for rolling element bearings 104
4.2.1 Modelling 105
4.2.2 Parameter estimation 106
4.2.3 Fault detection 107
4.2.4 Experimental validation 109
4.2.5 Concluding remarks 114
4.3 Summary and concluding remarks 115

[1] © 2021 IEEE. Reprinted with permission from Othmane, A., & Rudolph, J. Data and computation efficient model-based fault detection for rolling element bearings using numerical differentiation. In Proc. of the 5th International Conf. on Control and Fault-Tolerant Systems (p. 163-168). Saint-Raphaël, France: IEEE. doi: 10.1109/SysTol52990.2021.9595974 (2021).

Figure 4.1: Experimental setup of a plate which may move in the field of four electromagnets with a table tennis ball. The vertical position of the plate is captured by four sensors. (Photo: Saarland University, Oliver Dietze)

4.1 Model-based collision detection on a magnetically supported plate

The use of Jacobi and Laguerre diffs. for the detection of impulsive forces is first demonstrated on an experimental setup shown in Fig. 4.1. This setup has already been used in Kiltz (2017); Kiltz et al. (2013); Kiltz et al. (2014, 2012); Kiltz and Rudolph (2013). It consists of a $4.7\,\mathrm{kg}$ rectangular aluminium plate with four iron profiles at its corners. The plate can be lifted by forces generated by four electromagnets mounted to the frame above the plate. Sensors measure the vertical displacement between the plate and the frame. The currents through the magnet coils are driven by current controllers. The controller for the position has been discussed in Kiltz et al. (2014, 2012).

In this section the following scenario is considered. While the plate is hovering at a constant vertical position equal to $-4.5\,\mathrm{mm}$, a $2.7\,\mathrm{g}$ table tennis ball is falling on it. The task is to first detect a collision and then to estimate the unknown collision time t_c, i.e., the time instant when the ball hits the plate, using only the displacement measurement and the known currents. The difficulties here are the significant difference in the mass between the ball and the plate and the highly disturbed displacement measurement. In fact, an (audible) mechanical eigenmode of the plate corrupts the measurements. False alarms due to disturbances and unmodelled dynamics must be prevented.

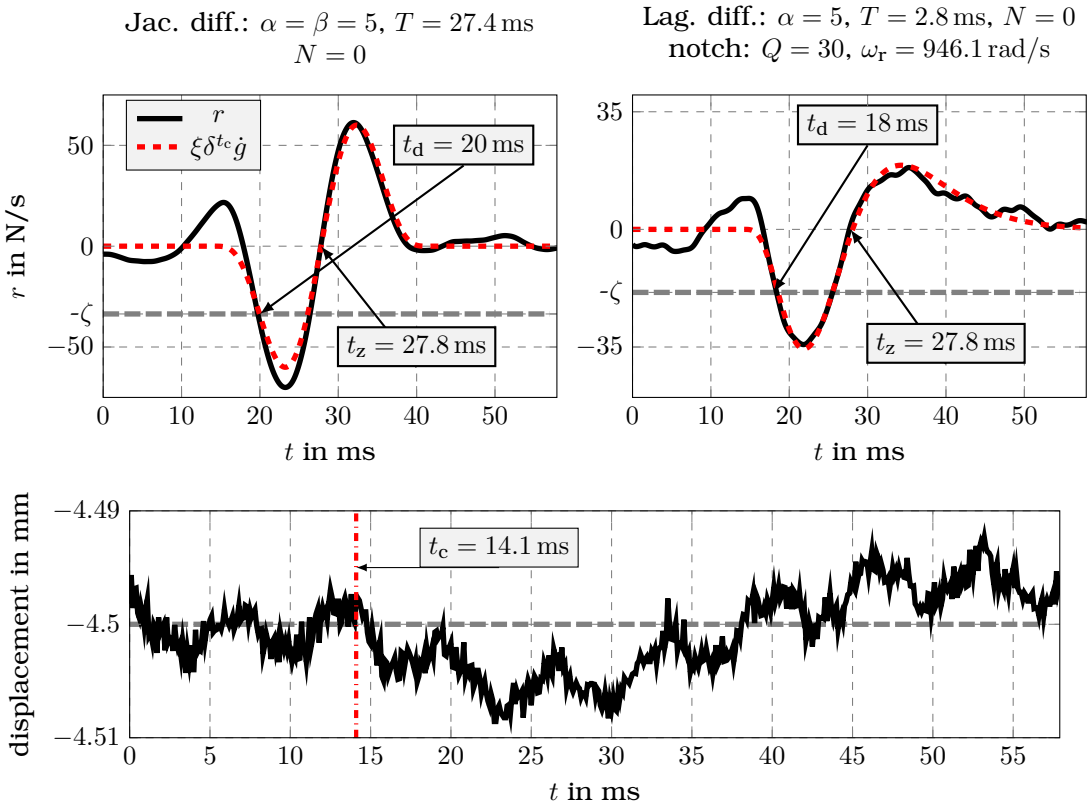


Figure 4.2: Collision detection using Jacobi and Laguerre diffs. at a specific hovering position of the plate equal to $y = -4.5\,\mathrm{mm}$ and a falling height of the ball equal to $3\,\mathrm{cm}$: The detection time and the time instant where the residual r from (4.2) crosses the zero line after the maximum excitation are denoted by t_d and t_z, respectively. The estimated collision time is t_c.

4.1.1 Collision detection and collision time estimation

The vertical motion of the plate may be described as in Kiltz (2017); Kiltz and Rudolph (2013) by the differential equation

$$M\ddot{y} + M\Gamma - F = -C + \nu, \tag{4.1a}$$

with $y : \mathbb{R} \to \mathbb{R}$ the measured vertical displacement of the plate, $F : \mathbb{R} \to \mathbb{R}$ the vertical force generated by the magnets, $C : \mathbb{R} \to \mathbb{R}$ the contact force between the ball and the plate, and $\nu : \mathbb{R} \to \mathbb{R}$ model uncertainties, that are assumed to be slowly varying. Moreover, $M = 4.691\,\mathrm{kg}$ is the mass of the plate and $\Gamma = 9.8091\,\mathrm{m/s^2}$ is the gravitational acceleration. The contact force derived in Kiltz (2017) can be modelled as

$$C(t) = \xi\delta(t - t_\mathrm{c}), \quad \xi = m(1 + \mathrm{k})\sqrt{2\Gamma H}, \tag{4.1b}$$

where the weight[2] ξ of the Dirac impulse δ depends on the mass $m = 2.7\,\mathrm{g}$, the restitution coefficient $\mathrm{k} = 0.9314$, and the falling height H of the ball. Due to the use of the Dirac impulse in the model, functions are identified with their induced distributions in the sense of Schwartz (see, e.g., Schwartz (1966)) in the sequel.

[2]The weight ξ in (4.1b) corresponds to the change in momentum of the ball as a result of the collision (see (Kiltz, 2017, Sec. 4.1) for more details)

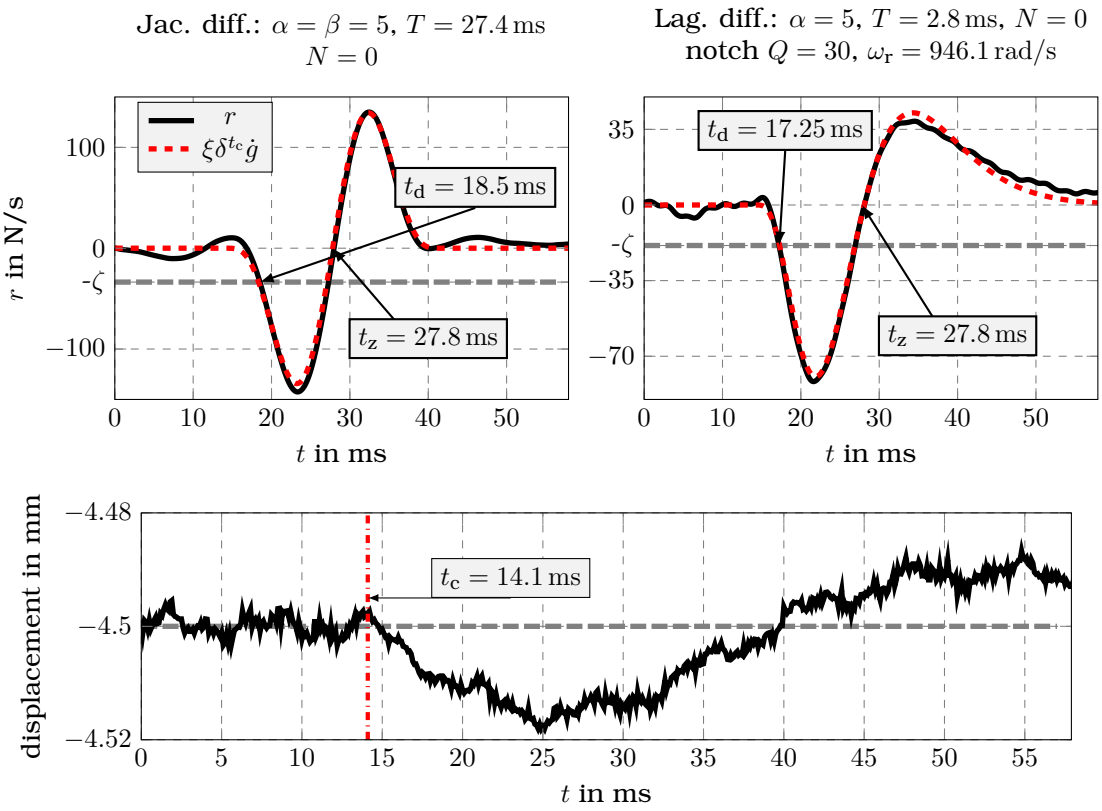


Figure 4.3: Collision detection using Jacobi and Laguerre diffs. at a specific hovering position of the plate equal to $y = -4.5\,\mathrm{mm}$ and a falling height of the ball equal to $15\,\mathrm{cm}$: The detection time and the time instant where the residual r from (4.2) crosses the zero line after the maximum excitation are denoted by t_d and t_z, respectively. The estimated collision time is t_c.

Let g be an orthogonal differentiator (orth. diff.). A signal $r : \mathbb{R} \to \mathbb{R}$ that indicates a collision has been derived in Kiltz (2017); Kiltz and Rudolph (2013) as[3]

$$r = Mg^{(3)} \star y - \dot{g} \star F. \tag{4.2}$$

The latter equation can be recovered by differentiating (4.1a) with respect to time in order to eliminate slowly varying model uncertainties. The signal r, which in the sequel is called the residual, can be computed using solely the known signals of the displacement and the applied force. Using (4.1a) yields

$$r = -\xi\delta^{t_\mathrm{c}}\dot{g} + g \star \dot{\nu}, \tag{4.3}$$

where $\delta^{t_\mathrm{c}}\dot{g}$ corresponds to the function $\dot{g}$ delayed by the delay t_c. Thus, a reliable detection of collisions can be achieved if the amplitude of $\xi\delta^{t_\mathrm{c}}\dot{g}$ is much larger than that of $g \star \dot{\nu}$.

In the absence of a collision the residual corresponds to the zero-mean signal $\dot{g} \star \nu$. To detect collisions it is then sufficient to compare the residual to a constant threshold the value of which should be chosen large enough to prevent false alarms. Consequently, only a ball falling from a minimal height H_min can be detected. Its value depends on the choice of the diff. as will be discussed in the next subsection.

[3]The convolution of two functions f and g is written $f \star g$.

When a collision occurs, the impulse response of $\dot{g}$ scaled by ξ stimulates the residual, as given in (4.3). When the diff. is properly parametrised a ball falling from low heights can be detected as demonstrated in Fig. 4.2 for a height of $3\,\mathrm{cm}$ using Jacobi and Laguerre diffs. When the residual crosses the threshold, a collision is detected. The crossing time instant denoted by t_d is called detection time in the sequel and depends on the chosen diff. As first observed in Kiltz (2017), the detection time depends highly on the falling height of the ball (see, e.g., Figs. 4.2, 4.3 and 4.9). Therefore, the detection time generally does not allow an exact reconstruction of the collision time. On the other hand, the collision time, denoted by t_c, can be estimated with good precision from the zero-crossing that follows the maximum excitation of the collision indicator. This is also illustrated in Figs. 4.2 and 4.3. From the figures it is clear that the measurement of the displacement alone is insufficient for a reliable estimation of the collision time as opposed to the derived residual.

Differentiator parametrisations allowing the detection of collisions for low falling heights and their implementations are now discussed. Jacobi and Laguerre diffs. are used, and the estimation quality, detection times, and the implementation burdens are compared.

4.1.2 Parametrisation of differentiators

If the detection threshold is chosen as $\zeta = \zeta_\mathrm{max}\sigma_{\dot{g}\star\nu}$, with $\sigma_{\dot{g}\star\nu}$ the standard deviation of the signal $\dot{g}\star\nu$ in the collision-free case, then the minimal falling weight required for the residual to cross the threshold is

$$\xi_\mathrm{min} \approx \frac{\zeta_\mathrm{max}\sigma_{\dot{g}\star\nu}}{\max_\tau |\dot{g}(\tau)|}.$$

From (Kiltz, 2017, Ch. 4), this weight corresponds to the minimal detectable falling height of

$$H_\mathrm{min} \approx \frac{1}{2\Gamma}\left(\frac{\zeta_\mathrm{max}\sigma_{\dot{g}\star\nu}}{m(1+\mathrm{k})\max_\tau |\dot{g}(\tau)|}\right)^2.$$

The diffs. shall be parametrised to minimise the minimal detection height. In the following $\zeta_\mathrm{max} = 5$, which corresponds to a false alarm probability less than $0.6\cdot 10^{-4}\%$ if the collision-free residual is assumed to be drawn from a normal distribution. In Fig. 4.4 the cumulative distribution function (CDF) of a normal distribution is compared to that of the measurement data for residuals computed using the Jac. diff. parametrised in Kiltz (2017) and the approach with two different Lag. diff. This shows that the choice of the threshold is reasonable to have a negligible false alarm probability.

In the absence of collisions the relation $\dot{\nu} = My^{(3)} - \dot{F}$ holds. Hence, an estimate of the discrete Fourier transform $\mathcal{D}\{\dot{\nu}\}(k) = \sum_{n=0}^{L_\mathrm{m}-1}\dot{\nu}(nt_\mathrm{s})\mathrm{e}^{-\iota 2\pi nk/L_\mathrm{m}}$ of $\dot{\nu}$ is

$$\hat{\mathcal{D}}\{\dot{\nu}\}(k) = -\iota\omega_k\left(\omega_k^2\mathcal{D}\{y\}(k) + \mathcal{D}\{F\}(k)\right) \tag{4.4}$$

with $\omega_k = \frac{2\pi k}{L_\mathrm{m}t_\mathrm{s}}$, $t_\mathrm{s} = 0.1\,\mathrm{ms}$ the sampling period, and L the number of measurement samples (see, e.g., (Oppenheim & Schafer, 1989, Ch. 8)). Fig. 4.5 shows an estimate of the amplitude spectrum of $\dot{\nu}$, where a particularly dominant peak arising from the previously mentioned audible mechanical eigenmode at $\omega_\mathrm{r} = 946.1\,\mathrm{rad/s}$ can be seen. Thus, to be able to detect collisions from low heights a suppression of this disturbing component is necessary.

Jacobi differentiator

In (Kiltz, 2017, Ch. 4), it has been shown that for a Jac. diff. with $\alpha = \beta = 5$, $N = 0$, and $T = 27.4\,\mathrm{ms}$ a minimal detected falling height of $H_\mathrm{min} = 0.94\,\mathrm{cm}$ with a detection time $t_\mathrm{d} =$

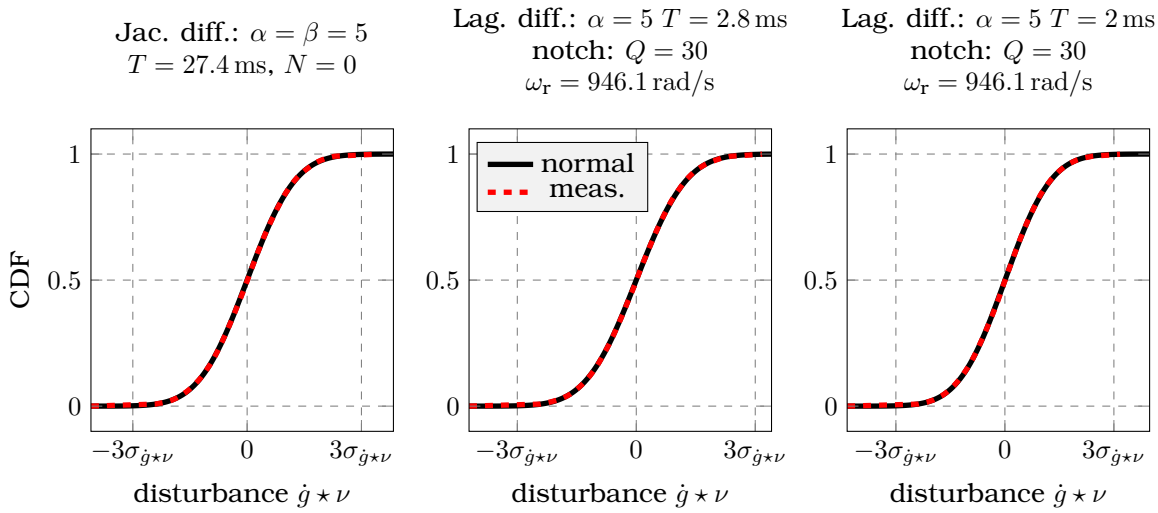


Figure 4.4: CDFs for the residuals in the collision-free case using measurements of the vertical displacement at a hovering position $y = -4.5\,\mathrm{mm}$.

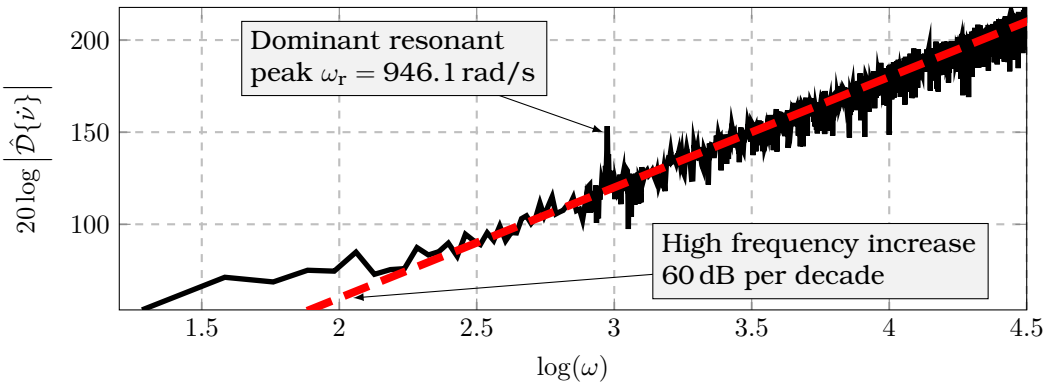


Figure 4.5: Estimated amplitude spectrum from (4.4) of the disturbance $\dot{\nu}$ based on $L_\mathrm{m} = 2^{17}$ measurements of the vertical displacement at a hovering position $y = -4.5\,\mathrm{mm}$.

$9.1\,\mathrm{ms}$ can be achieved. It has also been shown that for this configuration, the resulting filter approximates a matched filter (see, e.g., Turin (1960) for an introduction to matched filters). Fig. 4.8 shows the amplitude spectrum of this diff. It can be seen that its second local minima, the second zero of the Bessel function (see Section 3.1.3 where the zeros of the amplitude spectrum are discussed), corresponds to the dominant resonant peak in the amplitude spectrum of the disturbance. Thus, the effects of mechanical vibrations are exactly annihilated. In the previous discussion on the estimation of the collision time, the zero-crossing of the residual after its maximum excitation has been identified as a mean for the estimation of t_c. The impulse response crosses the zero line exactly at the centre of the window of the diff. given by $T/2 = 13.7\,\mathrm{ms}$. The reader is referred to (Kiltz, 2017, Ch. 4) for a detailed discussion on the parametrisation.

Laguerre differentiator

Since the amplitude spectrum of the Lag. diffs. does not exhibit zeros, an exact annihilation of the effects of the mechanical vibrations is impossible using only a diff. To this end, the residual

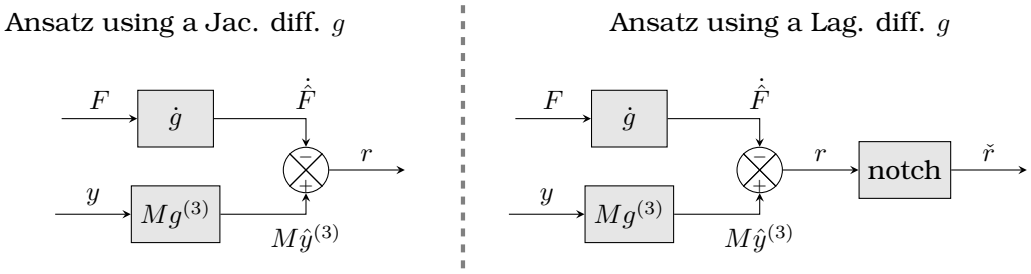


Figure 4.6: Comparison of the signal flow diagrams for the residual generation using the approaches with the Jac. diff. and the Lag. diff.

r is further filtered by a second-order notch filter. The resulting signal is denoted by $\check{r}$. Fig. 4.6 summarises the signal flow diagram of the approach in a more comprehensive representation and compares it with the previous approach based on the Jac. diff. The transfer function of the notch filter follows from (Tietze et al., 2008, Sec. 13.8) and is

$$h_r^{\check{r}}(\iota\omega) = \frac{-\left(\frac{\omega}{\omega_\mathrm{r}}\right)^2 + 1}{-\left(\frac{\omega}{\omega_\mathrm{r}}\right)^2 + \iota\frac{\omega}{\omega_\mathrm{r}}\frac{1}{Q} + 1}$$

where the parameter Q is called the quality factor. The larger Q the steeper the gain falls off in the vicinity of the dominant resonant peak ω_r of the mechanical vibrations and the more narrow the stopband is.

In the presence of a collision, the new residual $\check{r}$ is a linear combination of the impulse response of the cascade of the notch filter and the first-order derivative of the Lag. diff. and the influence of the disturbances. In the previous discussion on the estimation of the collision time t_c, the zero-crossing of the residual after its maximum excitation has been identified as a mean for the estimation of t_c. Fig. 4.7 shows the impulse responses of the cascade and the first-order derivative of the diff. for $Q = 30$ and two Lag. diffs. with $N = 0$, $\alpha = 5$, and two different values for T. It can be seen that while the cascade has a different impulse response than the first-order derivative of the diff., the differences in the zero-crossing are negligible and are less than a sampling period. Thus, the zero-crossing of the first-order derivative of the Lag. diff., whose location at $T\alpha$ is precisely known, can be used to estimate the collision time. Decreasing the parameter Q of the notch filter, which enlarges the stopband, and keeping the parameters of the Lag. diff. constant will affect the impulse response. The zero of $\dot{g}$ is no longer a good estimate for the collision time.

Fig. 4.8 shows the amplitude spectra of the cascade of filters with different parameters. It can be seen that the dominant resonance frequency is nicely annihilated. For both parametrisations of the Lag. diff. the amplitude is higher or equal to that of the Jac. diff. for high frequencies. For low frequencies, the amplitude of the first is lower than that of the second one. This has already been observed in Section 3.3.

The parametrisation with $T = 2.8\,\mathrm{ms}$ has been used for the collision detection in the two examples from Figs. 4.2 and 4.3. It can be seen that the detection times are $t_\mathrm{d} = 18\,\mathrm{ms}$ and $t_\mathrm{d} = 17.25\,\mathrm{ms}$, for the collision from $3\,\mathrm{cm}$ and $15\,\mathrm{cm}$, respectively. For both collisions the detection time is lower when the Lag. diff. is used compared to those for the Jac. diff. For the collision from $3\,\mathrm{cm}$ the collision is detected $2\,\mathrm{ms}$ faster, which corresponds to 20 sampling periods and a decrease of approximately $10\,\%$. Both approaches estimate the same collision

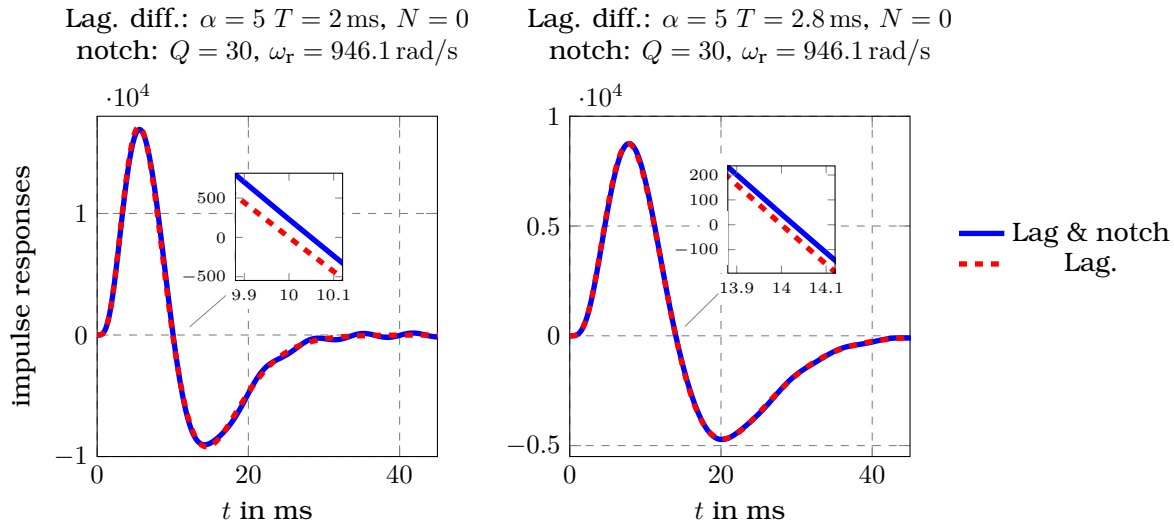


Figure 4.7: Impulse responses of the cascade of a first-order derivative of Lag. diff. and a notch filter and the first-order derivative of the diff. The parameters $Q = 30$ and $\omega_\mathrm{r} = 946.1\,\mathrm{rad/s}$ are chosen for the notch filter and two Lag. diffs. with $N = 0$, $\alpha = 5$, and two different values for T are considered.

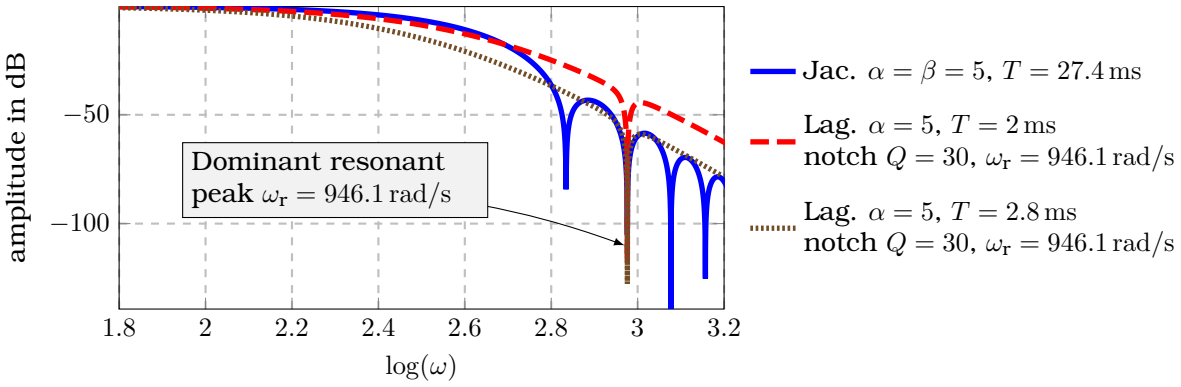


Figure 4.8: Amplitude spectra of the used diffs. for the collision detection.

time.

The faster detection observed in these two examples is confirmed by Fig. 4.9 where the detection time is illustrated as a function of the falling height computed using collision free measurements. The Lag. diff. with $T = 2.8\,\mathrm{ms}$ has a minimum detectable falling height of $0.83\,\mathrm{cm}$ compared to the Jac. diff. with $0.94\,\mathrm{cm}$. This corresponds to a decrease for H_min of approximately $11.7\,\%$. The detection time for the approach with the Lag. diff. is significantly lower than the one with the Jac. diff. Decreasing the parameter T of the Lag. diff. to $2\,\mathrm{ms}$ increases the minimal detectable falling height to $H_\mathrm{min} = 2.3\,\mathrm{cm}$. Thus, decreasing T by $40\,\%$ increases H_min by $177\,\%$.

4.1.3 Implementation of detection algorithms

The discussions started in Section 3.4 comparing the computational burden and the storage requirements for the simple examples are now extended for this concrete case study.

Jacobi differentiator

The window length of the Jac. diff. g with the previously mentioned parametrisation equals 274 sampling periods. Thus, computing the first-order derivative of the force F using $\dot{g}$ and the third-order derivative of the displacement using $g^{(3)}$ requires 1094 arithmetic operations in each sampling step, if the mid-point rule is used for the discretisation. For the implementation, 548 filter coefficients and past measurements have to be stored. In contrast, the recursive implementation schemes briefly discussed in Section 3.2.3, has been used in (Kiltz et al., 2014, Sec. 4.3) to reduce the required arithmetic operations by $88.1\,\%$ and the storage requirements by $84.3\,\%$. Thus, 130 operations are required in each step, and only 86 filter coefficients have to be stored. The recursive approximation of $\dot{F}$ and $y^{(3)}$ require an implementation in a 26 bit and 33 bit fixed point arithmetic system, respectively.

Laguerre differentiator

Discretising the Lag. diffs. for the approximation of $\dot{F}$ and $y^{(3)}$ as in Section 3.2.2 by an infinite-impulse response (IIR) filters using the bilinear transform yields the recursive scheme (3.19), i.e.,

$$\dot{\hat{F}}_k = \sum_{n=1}^{6} p_n \dot{\hat{F}}_{k-n} + \sum_{n=0}^{6} q_{n,F} F_{k-n}, \quad \hat{y}_k^{(3)} = \sum_{n=1}^{6} p_n \hat{y}_{k-n}^{(3)} + \sum_{n=0}^{6} q_{n,y} y_{k-n},$$

with

$$\begin{aligned}
p_1 &= -\frac{-384\,T^6 - 768\,T^5 t_\mathrm{s} - 480\,T^4 {t_\mathrm{s}}^2 + 120\,T^2 {t_\mathrm{s}}^4 + 48\,T {t_\mathrm{s}}^5 + 6\,{t_\mathrm{s}}^6}{p_0},\\
p_2 &= -\frac{960\,T^6 + 960\,T^5 t_\mathrm{s} - 240\,T^4 {t_\mathrm{s}}^2 - 480\,T^3 {t_\mathrm{s}}^3 - 60\,T^2 {t_\mathrm{s}}^4 + 60\,T {t_\mathrm{s}}^5 + 15\,{t_\mathrm{s}}^6}{p_0},\\
p_3 &= -\frac{-1280\,T^6 + 960\,T^4 {t_\mathrm{s}}^2 - 240\,T^2 {t_\mathrm{s}}^4 + 20\,{t_\mathrm{s}}^6}{p_0},\\
p_4 &= -\frac{960\,T^6 - 960\,T^5 t_\mathrm{s} - 240\,T^4 {t_\mathrm{s}}^2 + 480\,T^3 {t_\mathrm{s}}^3 - 60\,T^2 {t_\mathrm{s}}^4 - 60\,T {t_\mathrm{s}}^5 + 15\,{t_\mathrm{s}}^6}{p_0},\\
p_5 &= -\frac{-384\,T^6 + 768\,T^5 t_\mathrm{s} - 480\,T^4 {t_\mathrm{s}}^2 + 120\,T^2 {t_\mathrm{s}}^4 - 48\,T {t_\mathrm{s}}^5 + 6\,{t_\mathrm{s}}^6}{p_0},\\
p_6 &= -\frac{64\,T^6 - 192\,T^5 t_\mathrm{s} + 240\,T^4 {t_\mathrm{s}}^2 - 160\,T^3 {t_\mathrm{s}}^3 + 60\,T^2 {t_\mathrm{s}}^4 - 12\,T {t_\mathrm{s}}^5 + {t_\mathrm{s}}^6}{p_0},\\
p_0 &= 64\,T^6 + 192\,T^5 t_\mathrm{s} + 240\,T^4 {t_\mathrm{s}}^2 + 160\,T^3 {t_\mathrm{s}}^3 + 60\,T^2 {t_\mathrm{s}}^4 + 12\,T {t_\mathrm{s}}^5 + {t_\mathrm{s}}^6,
\end{aligned}$$

and

$$\begin{aligned}
q_{0,F} = \frac{2\,{t_\mathrm{s}}^5}{p_0},\ q_{1,F} = \frac{8\,{t_\mathrm{s}}^5}{p_0},\ q_{2,F} = \frac{10\,{t_\mathrm{s}}^5}{p_0},\ q_{3,F} = 0,\ q_{4,F} = \frac{-10\,{t_\mathrm{s}}^5}{p_0},\ q_{5,F} = \frac{-8\,{t_\mathrm{s}}^5}{p_0},\ q_{6,F} = \frac{-2\,{t_\mathrm{s}}^5}{p_0},\\
q_{0,y} = \frac{8\,{t_\mathrm{s}}^3}{p_0},\ q_{2,y} = -24\,\frac{{t_\mathrm{s}}^3}{p_0},\ q_{4,y} = 24\,\frac{{t_\mathrm{s}}^3}{p_0},\ q_{6,y} = -8\,\frac{{t_\mathrm{s}}^3}{p_0},\ q_{1,y} = q_{3,y} = q_{5,y} = 0.
\end{aligned}$$

Thus, the approximation of the two derivatives requires storing 16 filter coefficients, 12 past filter outputs (six for each) and 12 measurement samples (six for each). In very sampling period, 42 arithmetic operations have to be performed to compute the residual r.

For the annihilation of the disturbance stemming from the audible mechanical eigenmode of the plate, the mentioned notch filter can be discretised using the bilinear transform with

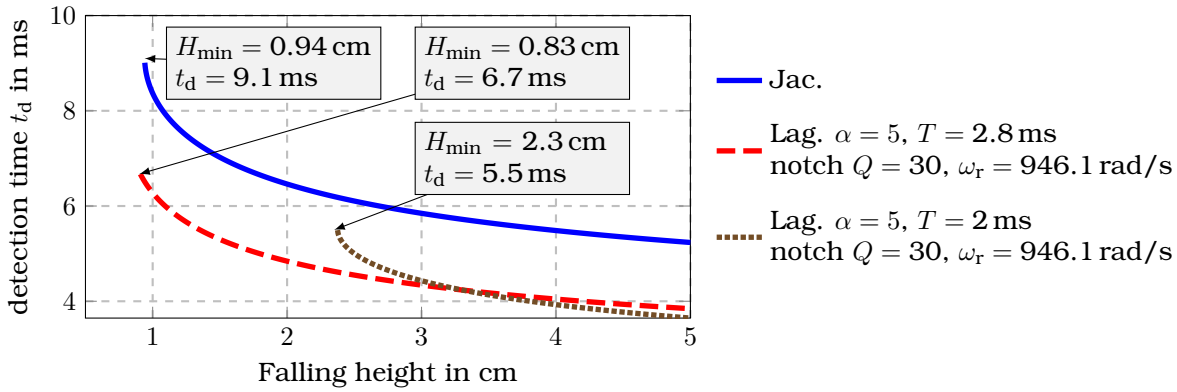


Figure 4.9: Detection time t_d for the nominal time-evolution of the collision indicator $\xi\delta^{t_\mathrm{c}}\dot{g}$ as a function of the falling height of the ball and different diffs. computed using collision free measurements.

frequency prewarping (see, e.g., (Oppenheim & Schafer, 1975, Ch. 5) and (Franklin et al., 1998, Ch. 6)). The prewarp frequency is ω_r. Then, the residual $\check{r}$ can be approximated as

$$\check{r}_k = \sum_{n=1}^{2} p_{n,\mathrm{N}}\check{r}_{k-n} + \sum_{n=0}^{2} q_{n,\mathrm{N}} r_{k-n}$$

with

$$p_{1,\mathrm{N}} = \frac{4\cos^2\left(\frac{\omega_\mathrm{r} t_\mathrm{s}}{2}\right)Q - 2\,Q}{p_{0,\mathrm{N}}}, \quad p_{2,\mathrm{N}} = \frac{\sin\left(\frac{\omega_\mathrm{r} t_\mathrm{s}}{2}\right)\cos\left(\frac{\omega_\mathrm{r} t_\mathrm{s}}{2}\right) - Q}{p_{0,\mathrm{N}}}, \quad p_{0,\mathrm{N}} = -\sin\left(\frac{\omega_\mathrm{r} t_\mathrm{s}}{2}\right)\cos\left(\frac{\omega_\mathrm{r} t_\mathrm{s}}{2}\right) - Q,$$

$$q_{0,\mathrm{N}} = \frac{-Q}{p_{0,\mathrm{N}}}, \quad q_{1,\mathrm{N}} = \frac{4\cos^2\left(\frac{\omega_\mathrm{r} t_\mathrm{s}}{2}\right)Q - 2\,Q}{p_{0,\mathrm{N}}}, \quad q_{2,\mathrm{N}} = \frac{-Q}{p_{0,\mathrm{N}}}.$$

This implementation of the notch filter requires storing five filter coefficients, two past filter outputs, and two measurement samples of the residual r. In total nine arithmetic operations are required in each sampling step.

To conclude, computing the residual $\check{r}$ using the Lag. diff. and the notch filter requires 51 arithmetic operations in each sampling step. In total 21 filter coefficients, 14 past filter outputs, and 14 measurement samples (including the two samples of r) have to be stored. Comparing this implementation with the non-recursive implementation of the approach using the Jac. diff., the number of arithmetic operations is reduced by $95.33\,\%$. The method also decreases the number of filter coefficients to be stored by $96.17\,\%$. When the approach is compared to the recursive implementation from Kiltz (2017), the computational burden and the number of filter coefficients to be stored are reduced by $60.77\,\%$ and $75.58\,\%$, respectively.

4.1.4 Concluding remarks

The previous analyses have confirmed the results in Kiltz (2017); Kiltz and Rudolph (2013) showing that Jac. diff. can be used for the detection of collisions of a table tennis ball on a plate using only the displacement measurement of the plate and the applied magnetic forces. The unique properties of the Jac. diff. have enabled the annihilation of the disturbance stemming from the audible mechanical eigenmode of the plate. Then, the problem has also been solved using a Lag. diff. together with a notch filter. It has been shown that the approach

yields a lower minimal detectable falling height with a lower detection time. Both approaches estimate the same collision time that cannot be estimated using only the disturbed vertical displacement measurement of the plate. This new approach also significantly reduces the computational burden and the storage requirements when the recursive implementation of the filters is used.

4.2 Model-based fault detection for rolling element bearings

Rolling element bearings are of paramount importance to almost all forms of rotating machinery. Their failures are among the foremost causes of breakdowns in machines and may be catastrophic and result in costly downtime. In Yung and Bonnett (2004), it is claimed that at least 50% of electric motor failures start as bearing failures. According to de Azevedo, Araújo, and Bouchonneau (2016), vast quantities of failures of wind turbines are caused by bearing failures. To prevent catastrophic damages and costly downtime, early detection of failures is thus crucial.

Bearing faults can be classified into single, multiple, or distributed point faults. The single point defects are single localised defects on an intact bearing surface and may be cracks, pits, or holes on the inner race, outer race, roller elements, or cage. These defects generate characteristic frequencies that appear in the vibration or acoustic emission signals and depend on the bearing geometry as discussed in McFadden and Smith (1984) and Smith and Randall (2015). A characteristic frequency can be associated with every single point defect of the components of the bearing. Multiple single point defects can appear simultaneously, and the spectra of the vibrations differ from those observed in the presence of a single defect as analysed in McFadden and Smith (1985). The analysis and the detection of these failures are thus challenging. Finally, distributed point faults or generalised roughness faults are a deterioration of the bearing over a large bearing surface area. As discussed in Stack, Habetler, and Harley (2004) and Nectoux et al. (2012), these faults do not excite any of the characteristic fault frequencies, and their early detection may be complicated.

Different bearing condition monitoring approaches have been discussed in the literature. For several years signal processing methods (Betta, Liguori, Paolillo, and Pietrosanto (2002); C. Li, Sanchez, Zurita, Cerrada Lozada, and Cabrera (2016); Nikolaou and Antoniadis (2002); Smith and Randall (2015)) have been used. Machine learning methods, such as artificial neural networks and support vector machine methods, recently attracted much attention. In Daga et al. (2019) Principal Component Analysis and Linear Discriminant Analysis are used. An extensive review of data-driven methods can be found in Cerrada et al. (2018). A model-based approach for fault detection using linear observers is proposed in Qian (2019).

In this work, a novel model-based approach for detecting failures based on efficient and robust numerical differentiation is introduced. A simplified model for the vibrations is proposed. The parameters may be estimated from fault-free acceleration measurements using Jac. diffs. derived in Mboup et al. (2009) and analysed in Kiltz (2017). A detailed analysis of the estimation errors is carried out in Othmane, Rudolph, and Mounier (2020). This model is then used to design signals, called residuals, sensitive only to faults. They are easy to compute, interpret and implement. The time evolution of the residuals is proportional to the force stemming from the faults. An approach for a systematic derivation of a failure detection threshold is proposed using the properties of Jac. diffs. This threshold can be designed to

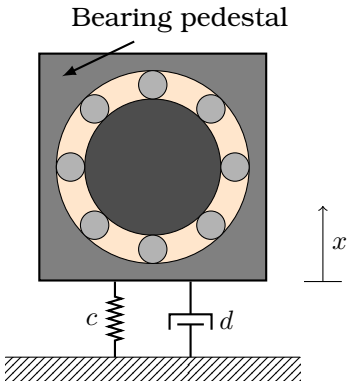


Figure 4.10: Simplified model of a rolling element bearing.

meet a desired probability of false alarms.

This section is structured as follows. The physically motivated model for the vibrations is discussed in Section 4.2.1. The estimation of the model parameters and the diff. parametrisation is explained in Section 4.2.2. An approach for the computation of a failure detection threshold is proposed in Section 4.2.3. Finally, the approach is validated using different experimental[4] data sets in Section 4.2.4.

4.2.1 Modelling

In this section, a physically motivated model for the acceleration of the bearings is sought. The contact between the bearing pedestal and the other components is modelled as a lumped spring-mass system, where only the vibration of the outer race is considered, as shown in Fig. 4.10. For simplicity, the vibrations in only one direction are considered here. However, the extension of the approach is straightforward. As the rolling elements pass through a defect position, an impulse force $f : \mathbb{R} \to \mathbb{R}$ will be generated. Thus, the vibration produced by a local fault can be modelled as the response of the bearing system to impulses that occur at different time instants (McFadden and Smith (1984, 1985)). The fault force f is modelled as a sum of weighted Dirac impulses

$$f(t) = \sum_{i \in \mathbb{N}} \gamma_i \delta(t - t_{f_i}), \tag{4.5a}$$

that affect the bearing housing at times $t_{f_i} \geqslant 0$, $i \in \mathbb{N}$. In the latter equation $\delta : \mathbb{R} \to \mathbb{R}$ is the Dirac impulse and $\gamma_i \in \mathbb{R}$, with $i \in \mathbb{N}$, describe the impulse strengths. In the sequel, the explicit dependence of functions on time is omitted when no misinterpretation is possible.

[4]The experimental validation of the approach developed in this work would not have been possible without the use of the publicly available experimental data sets. The author is greatly indebted to the research centres providing them.

Remark 4.1: Shaft rotating with time-varying velocity

It is assumed for notational simplicity that the shaft is rotating with a constant velocity. The generalisation of the approach is straightforward by assuming the velocity to be known. The system parameters are then a function of the velocity. The experimental validation shows that a set of identified parameters can yield good results for intervals of velocities.

The displacement $t \mapsto x(t)$ of the bearing pedestal satisfies

$$m_\mathrm{b}\ddot{x} + d\dot{x} + cx = u_0 u + f + h, \tag{4.5b}$$

where $m_\mathrm{b} \in \mathbb{R}$ denotes the mass of the outer race and associated bearing pedestal, $d \in \mathbb{R}$ and $c \in \mathbb{R}$ the damping and stiffness of the outer race and pedestal, respectively, and $u_0 \in \mathbb{R}$ the input gain. The system is excited by the input u. Model uncertainties and forces acting on the system such as gravity or radial forces are summarised in $h : \mathbb{R} \to \mathbb{R}$. In the sequel it is assumed that

$$\ddot{h} \approx 0 \tag{4.5c}$$

and that the input u satisfies the differential equation

$$\ddot{u} + \omega_u^2 u = 0, \tag{4.5d}$$

i.e., u is a harmonic function with frequency ω_u. Due to the introduction of the Dirac impulse in the model, functions are identified with their induced distributions in the sense of Schwartz (see, e.g., Schwartz (1966)) in the following.

In practice, the acceleration $\ddot{x}$ is measured by vibration transducers mounted at the bearing pedestal. In the sequel the measurement is denoted by $\tilde{y}$ and is assumed to satisfy $\tilde{y} = y + \eta$, with $y = \ddot{x}$ and η an additive disturbance.

4.2.2 Parameter estimation

To estimate all unknown parameters of the model (4.5) an equation involving only known signals and their derivatives is required, i.e., the input u has to be eliminated and the acceleration y incorporated. Differentiating (4.5b) twice with respect to time and using the model of the input (4.5d) yields

$$m_\mathrm{b} x^{(4)} + d x^{(3)} + (c + m_\mathrm{b}\omega_u^2)\ddot{x} + \omega_u^2 d\dot{x} + \omega_u^2 cx = \omega_u^2 f + \ddot{f} + \omega_u^2 h + \ddot{h}.$$

By differentiating the latter equation twice and using the assumption (4.5c) the model uncertainties vanish and a formulation

$$y^{(4)} + p_3 y^{(3)} + p_2 \ddot{y} + p_1 \dot{y} + p_0 y = q_4 f^{(4)} + q_2 \ddot{f} \tag{4.6}$$

with the acceleration y and the fault force f is obtained. In the latter equation $p_3 = d/m_\mathrm{b}$, $p_2 = c/m_\mathrm{b} + \omega_u^2$, $p_1 = \omega_u^2 d/m_\mathrm{b}$, $p_0 = \omega_u^2 c/m_\mathrm{b}$, $q_4 = 1/m_\mathrm{b}$, and $q_2 = \omega_u^2/m_\mathrm{b}$. Equation (4.6) can now be used to identify the parameters p_i, $i = 0, \ldots, 3$. Since the fault f is unknown, the identification of q_i, $i \in \{2, 4\}$, is not necessary.

In the fault-free case

$$\sum_{i=0}^{3} p_i y^{(i)} = -y^{(4)} \tag{4.7}$$

holds and the Jac. diffs. can be used for the estimation of the parameters. These Jac. diffs. can be seen here as a special choice of modulation functions first introduced by Shinbrot in Shinbrot (1957). Let g be a Jac. diff. with the parameters α, β, T_g, ϑ and N. Convolving both sides of (4.7) with g yields $\sum_{i=0}^{3} p_i g \star y^{(i)} = -g \star y^{(4)}$. Assuming $\alpha, \beta > 3$, the differentiation can be transferred from y to g to get $\sum_{i=0}^{3} p_i \hat{y}^{(i)} = -\hat{y}^{(4)}$, with $\hat{y}^{(i)} = g^{(i)} \star y$. Using $N_m + 1$ discrete measurements $\tilde{y}(t_k)$, $k \in \{0, \dots, N_m\}$ of the acceleration y, the four unknown parameters can be estimated by solving the least squares optimisation problem

$$J_{N,T_g,\vartheta}^{(\alpha,\beta)} = \min_{p_0,\dots,p_3 \in \mathbb{R}} \sum_{k=0}^{N_m} \left(\sum_{i=0}^{3} p_i \hat{y}^{(i)}(t_k) + \hat{y}^{(4)}(t_k) \right)^2 .$$

For the parametrisation of the Jac. diffs. the results of the frequency-domain analysis and the tuning guidelines developed in Kiltz (2017); Kiltz and Rudolph (2013); Mboup and Riachy (2018); Othmane, Rudolph, and Mounier (2021a) can be used. In particular, the choice $N = 0$ and $\alpha = \beta$ depicts the best robustness with respect to disturbances. From Kiltz and Rudolph (2013) the filter window length can be computed from a desired cutoff frequency. The parameters α and β have to be chosen such that the frequency-domain properties of the Jac. diff. are preserved after the discretisation. For small time window lengths T_g, or similarly high cuttoff frequencies ω_c, the parameters α and β have to be chosen high enough, such that these properties are preserved after the discretisation.

4.2.3 Fault detection

Residual generation

The model (4.6) containing only the acceleration and the faults, can now be used to derive a residual signal by replacing the parameters with their identified values. The residual is zero in the absence of faults and strictly greater than zero otherwise. It is first assumed that the model (4.6) perfectly describes the system. This assumption is dropped in the next subsection.

Denote by $\hat{p}_i$, $i = 0, \dots, 3$, the identified parameters from the last section and consider the signal

$$\bar{r} = \hat{y}^{(4)} + \sum_{i=0}^{3} \hat{p}_i \hat{y}^i. \tag{4.8a}$$

In the absence of faults, $\bar{r}$ contains only measurement noise and errors stemming from the parameter estimation. These errors are analysed in detail in Othmane et al. (2020). In the presence of a fault f, the function $\bar{r}$ satisfies

$$\bar{r} \approx q_4 g \star f^{(4)} + q_2 g \star \ddot{f} = q_4 g^{(4)} \star f + q_2 \ddot{g} \star f.$$

Defining $\varphi = q_4 g^{(4)} + q_2 \ddot{g}$ and using the general definition of the fault in (4.5a), which corresponds to the force generated by rolling elements striking a defect, yields

$$\bar{r}(t) = \int_{-\infty}^{\infty} \varphi(\tau) f(t - \tau) \mathrm{d}\tau = \sum_{i \in \mathbb{N}} \gamma_i \varphi(t - t_{f_i}).$$

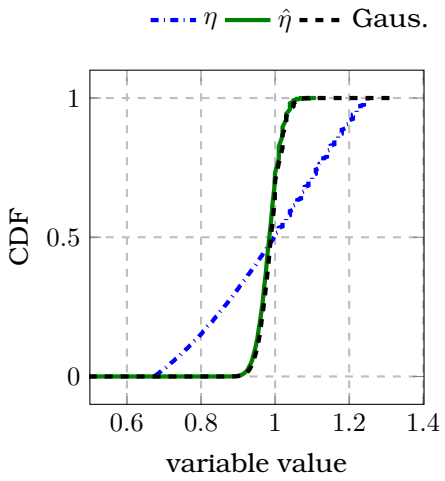


Figure 4.11: CDFs of a random variable $\eta = \epsilon^5 + \sqrt{(\epsilon+1)}$ with ϵ drawn uniformly from $[-0.5, 0.5]$, the filtered variable $\hat{\eta} = g^{(3,3)}_{0,T_g,\vartheta} \star \eta$, and a normal distribution fitted to $\hat{\eta}$. The parameters of the differentiator are $T_g = 0.045$, $N = 0$, and $\alpha = \beta = 3$.

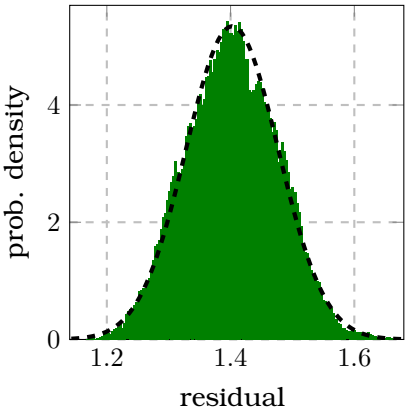


Figure 4.12: Probability distribution of the residual from (4.8) and a fitted Gaussian distribution with mean $\mu = 1.4\,\mathrm{m/s^4}$ and standard deviation $\sigma = 0.07\,\mathrm{m/s^4}$ for the fault-free system of the CWRU data set discussed in Section 4.2.4.

The signal $\bar{r}$ is now used to derive a residual that is proportional to the impulses strengths. Let $\mathcal{P} \subset \mathbb{R}$ be the domain of definition of the angle of the rotating shaft with respect to a reference and $\mathcal{I} \subset \mathbb{R}$ be the time domain of interest. Let $\phi : \mathcal{I} \to \mathcal{P}$ map every time instant to an angle. Assume ϕ is locally bijective and denote by ϕ^{-1} its inverse. From $\bar{r}$, a residual can be designed as

$$r(t) = \frac{1}{\bar{T}\kappa(t)} \int_{\phi(t)-\phi_1}^{\phi(t)} \left|\bar{r}\left(\phi^{-1}(\varphi)\right)\right| \mathrm{d}\varphi, \tag{4.8b}$$

with $\phi_1 \in \mathcal{P}$ and

$$\kappa(t) = \int_{\phi(t)-\phi_1}^{\phi(t)} \left| g^{(4)}(\phi^{-1}(\varphi)) + \tfrac{\hat{p}_1}{\hat{p}_3} \ddot{g}(\phi^{-1}(\varphi)) \right| \mathrm{d}\tau.$$

The residual (4.8b) can then be computed using (4.8a). For a better understanding of the residual (4.8), the special case of a shaft rotating with a constant velocity is considered. The residual then simplifies to

$$r(t) = \frac{1}{\bar{T}\kappa} \int_{t-\bar{T}}^{t} |\bar{r}(\tau)| \,\mathrm{d}\tau, \quad \mathbb{R} \ni \bar{T} > T_g. \tag{4.9}$$

For the interpretation of this residual, assume f consists only of one impulse, i.e., $f(t) = \gamma_0 \delta(t - t_{f_0})$, where $t_{f_0} > \bar{T}$. Then, the residual r from (4.8) reads

$$\begin{aligned} r(t) &= \tfrac{q_4}{\bar{T}\kappa} \int_{t-\bar{T}}^{t} \left| \gamma_0 \left(g^{(4)}(\tau - t_{f_0}) + \tfrac{q_2}{q_4} \ddot{g}(\tau - t_{f_0}) \right) \right| \mathrm{d}\tau \\ &\propto \begin{cases} |\gamma_0|, & \text{for} \quad t_{f_0} \in [t - \bar{T}, t - T_g], \\ 0, & \text{otherwise}, \end{cases} \end{aligned}$$

i.e., $r(t)$ is proportional to the intensity $|\gamma_0|$ of the impulse occurring at t_{f_0} for $t_{f_0} \in [t - \bar{T}, t]$ and zero otherwise. This can be generalised when the fault is a sequence of impulses as

initially defined in (4.5a). Assuming that the instants at which the impulses occur satisfy $t_{f_{i+1}} - t_{f_i} > T_g$, for $i \in \mathbb{N}$, the residual is proportional to the moving average of the intensities of the impulse strengths in $[t - \bar{T}, t - T_g]$.

Decision making

Threshold for fault detection Due to model uncertainties and measurement noise, the residual r in (4.8b) is not identically zero in the fault-free case. Thus, a strategy is required to detect faults reliably. Assume the residual r is a random variable drawn from a Gaussian distribution with mean value μ_r and standard deviation σ_r. Fig. 4.12 is an example illustrating this approximation using experimental data (the theoretical background is discussed in the following subsection). A κ_r-fold of the standard deviation σ_r can be interpreted as a threshold value which is not exceeded by the residual with a certain probability (e.g. for $\kappa_r = 3$ or $\kappa_r = 4$ the probability is 99.73% or 99.99%, respectively, that the residual does not exceed the threshold value $\kappa_r \sigma_r$). Thus, a Gaussian distribution can be fitted to the fault-free data, and by varying κ_r, the threshold can be computed to obtain a desired probability of false alarms.

Theoretical background Assume the acceleration $y = \ddot{x} = 0$, i.e., the measurement $\tilde{y} = \eta$ consists only of the additive disturbance η, which is now assumed to be a random variable. It has already been observed in Kiltz (2017) that the output of an algebraic Jac. diff. applied to a stationary random variable can be well approximated by a random variable drawn from a Gaussian distribution. Fig. 4.11 confirms this observation with a simulation example.

Consider now the residual r in (4.9) which corresponds to the moving average of $\bar{r}$ after its discretisation. Assume that $\bar{r}(t_i)$, $i \in \{0, 1, \ldots\}$, are independent identically distributed random variables drawn from a normal distribution. As depicted in Fig. 4.12, the probability density of r can be well approximated by a Gaussian distribution for growing $\bar{T}$.

From these two observations, it can be concluded that even when the measurement noise is not drawn from a Gaussian distribution, the residual can be approximated by one.

4.2.4 Experimental validation

The residual is parametrised such that the integration in (4.8) is performed over three revolutions. The threshold for the detection of faults introduced in the last section is parametrised for all experiments such that the probability of a false alarm is 0.01%, i.e., $\kappa_r = 4$. In the following, the residual is normalised to become unitless.

Case Western Reserve University data set

Test rig description The approach is first validated using the data set from the Case Western Reserve University available at CWRU (2015) and introduced in Smith and Randall (2015). An encoder and a torque transducer are mounted on a shaft driven by an electric motor. A dynamometer is used to apply torque on the shaft, which is supported by the test bearings, including drive end and fan end bearings. Faults of diameters of 7 mil, 14 mil, 21 mil, 28 mil, and 40 mil where $1\,\text{mil} = 0.001\,\text{inch}$, have been introduced manually to the inner ring, the outer ring, and the rolling elements. Accelerometers are mounted on the bearing housing to collect vibration data sampled at $12\,\text{kHz}$ and $48\,\text{kHz}$. In this work, the data sampled at $12\,\text{kHz}$ is used. Fault-free measurements for 4 different motor speeds ranging from $1730\,\text{rpm}$

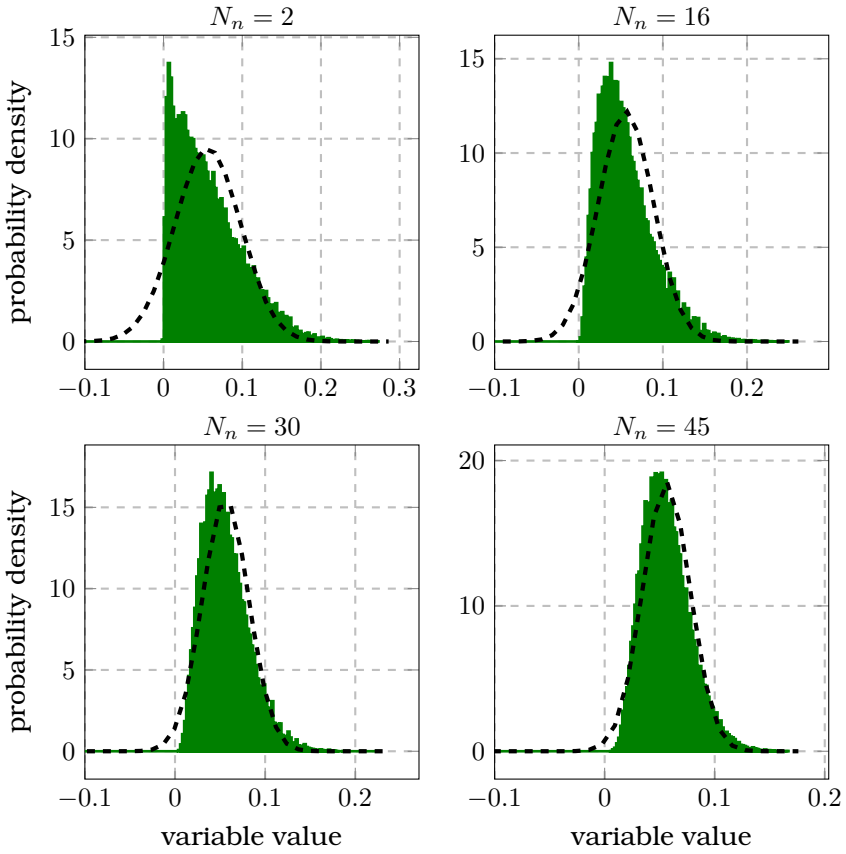


Figure 4.13: Evaluation of the residual (4.9) when the values $\bar{r}(t_i)$, $i \in \{0, 1, \ldots\}$, are identically distributed random variables drawn from a normal distribution and $\bar{T} = N_n t_s$.

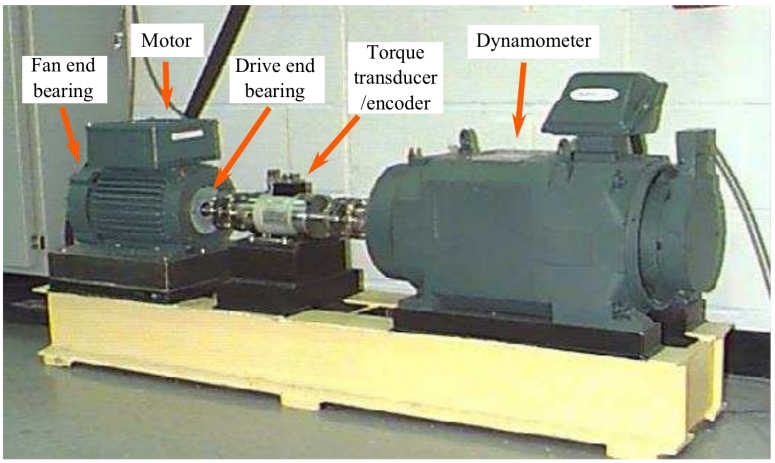


Figure 4.14: CWRU bearing test rig from Smith and Randall (2015).

to 1797 rpm are available. For illustration purposes, only the bearing of type 6205-2RSL JEM SKF is considered.

Experimental validation For the parameter estimation approach introduced in Section 4.2.2 the measurements for 0.25 s at the speeds 1730 rpm and 1797 rpm are used to identify one single set of parameters. This corresponds to only 6000 data points. Figure 4.12 depicts

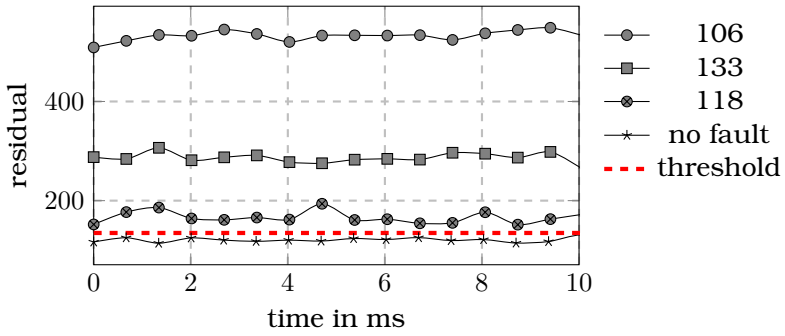


Figure 4.15: Time evolution of the residual of the experimental setup of the CWRU data set for the fault-free case and the experiments 106 (inner race fault), 133 (outer race centered fault), and 118 (ball fault). All defects have a diameter of $0.007\,\mathrm{inch}$.

the probability density of the residual in the fault-free case. Figure 4.15 depicts the time evolution of the residual for the fault-free case compared to the three different fault scenarios with the narrowest ($0.18\,\mathrm{mm}$) possible fault diameter: inner race, outer race, and roller element defect. The threshold for the detection of a fault is also given. All residuals are above the computed threshold. These results clearly show that a defect can be recognised and confirms the approach developed in this work. All, except 198 and 119, remaining faults in the experiments not considered here were correctly detected, even the defects at the rolling elements, which are certainly the most difficult to diagnose. These two datasets, namely 198 and 119, were not diagnosable with any of the methods applied in the initial work Smith and Randall (2015).

The Politecnico di Torino data set

Test rig description In Daga et al. (2009), the DIRG Lab in the Department of Mechanical and Aerospace Engineering at the Politecnico Di Torino provides measurements of a test bench described in detail in Daga et al. (2019). The test rig consists of a driven shaft supported by three bearings $\mathrm{B1}$, $\mathrm{B2}$, and $\mathrm{B3}$. The bearings $\mathrm{B1}$ and $\mathrm{B3}$ are identical. The outer ring of the bearing $\mathrm{B2}$ is linked to a precision sled moving in orthogonal direction with respect to the shaft. It produces a radial force acting on the shaft. The acceleration in all three spatial directions is measured. For more details the reader is referred to Daga et al. (2019). The monitoring of bearing $\mathrm{B1}$ is of interest, and in this work, only the acceleration of the tested bearing is used. The experiments at different speeds and loads are considered. Different types of localised defects have been manually incorporated using a Rockwell tool into the bearing $\mathrm{B1}$. The acceleration is measured at a sampling frequency $f_\mathrm{s} = 51.2\,\mathrm{kHz}$.

Experimental validation Since measurements of the acceleration in the three spatial directions are available, a model of the form (4.5) is assumed for each vibration axis, and three residuals r_i, $i \in \{1, 2, 3\}$, can be computed, one for each axis. These residuals can be combined to a single one $r = \sum_{i=1}^{3} r_i$. The threshold for the detection of a fault can be computed accordingly. Figure 4.17 depicts the results for different speed and load configurations. The top plot depicts the CDF of the residual in the fault-free case and a fitted Gaussian. The lower plot shows the residual time evolution and shows that the considered defects are correctly detected. However, the approach is not able to correctly detect all defects. For instance, the

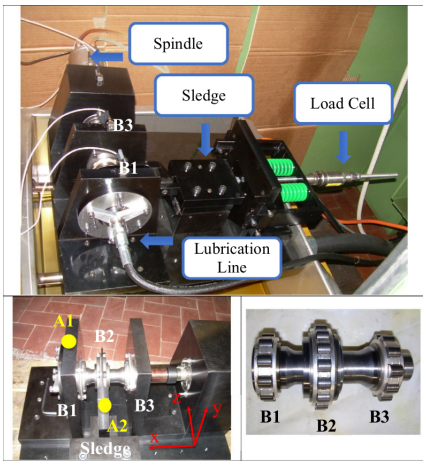


Figure 4.16: The test rig of the The Politecnico di Torino rolling bearing test data set from Daga et al. (2019).

defects (not depicted in the plot) of diameter $150\,\mu\mathrm{m}$ in the inner ring and on the roller could not be detected. The experimental setup can explain this: The bearings have been removed from the test bench and inserted back to incorporate the defects. Thus, the identified model does not correspond anymore to the monitored one. However, this problem is not a drawback of the approach since the bearing is not disassembled in industrial applications.

The PRONOSTIA data set

Test rig description The PRONOSTIA is an experimental platform described in Nectoux et al. (2012) and designed to provide experimental data characterising the degradation of bearings along their duration of operation. A synchronous motor drives a shaft supported by the bearing to be analysed through a gearbox and different couplings. A radial force is applied to the external ring of the test bearing. This force, the rotation speed of the shaft, and the acceleration of the bearing are measured. Two accelerometers are mounted on the bearing's external race, and the acceleration in the orthogonal directions is measured. For illustration purposes, only the data set Bearing1_1 is considered. It represents a run to failure experiment where the rotation speed of the shaft and the radial load are $1800\,\mathrm{rpm}$ and $4000\,\mathrm{N}$, respectively. For simplicity, only the vertical acceleration is considered for assessing the bearing health.

Experimental validation The parameters are identified and the threshold is computed using the methods discussed in the previous sections. In the top of Figure 4.19, the CDF of the residual in the fault-free case and a fitted Gaussian are depicted. This confirms the assumption that the residual can be modelled as a normally distributed random variable in the fault-free case. The figure also depicts the time evolution of the residual for the complete experiment and the standard deviation of the measured acceleration computed for each recording phase. The residual depicts an important jump happening approximately at the 50th minute of the test, followed by an increase in the residual value and a crossing of the threshold. After the jump the residual is increasing exponentially and an exponential model $t \mapsto c_0 \mathrm{e}^{c_1 t}$, with $c_0 = 924.57$ and $c_1 = 0.005/\mathrm{min}$, can be fitted to the data. This trend that can be seen at an early in the residual is not observed in the acceleration's standard deviation. This example exhibits again the effectiveness of the proposed approach.

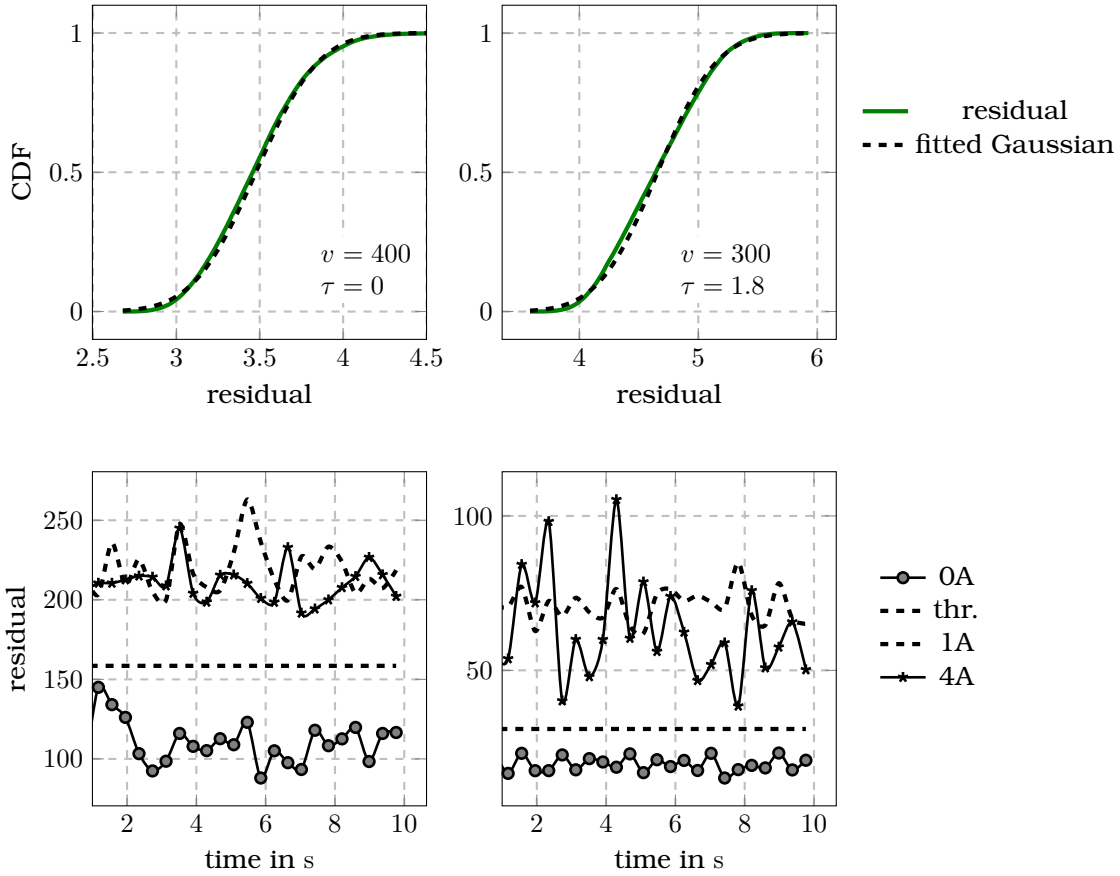


Figure 4.17: Analysis of the Politecnico di Torino data set. The rotation speed is denoted by v and is in kHz. The load is denoted by τ and is in kN. The signals are 0A for the fault-free case, 1A and 4A for the inner ring and roller faults of diameters $450\,\mu\text{m}$ and $450\,\mu\text{m}$, respectively.

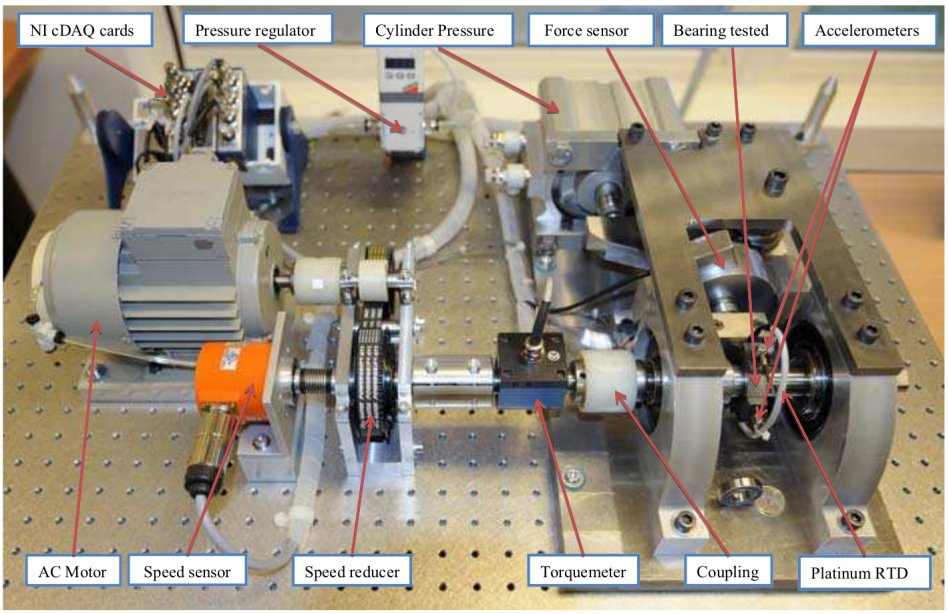


Figure 4.18: The test rig of the PRONOSTIA (Nectoux et al. (2012)) data set.

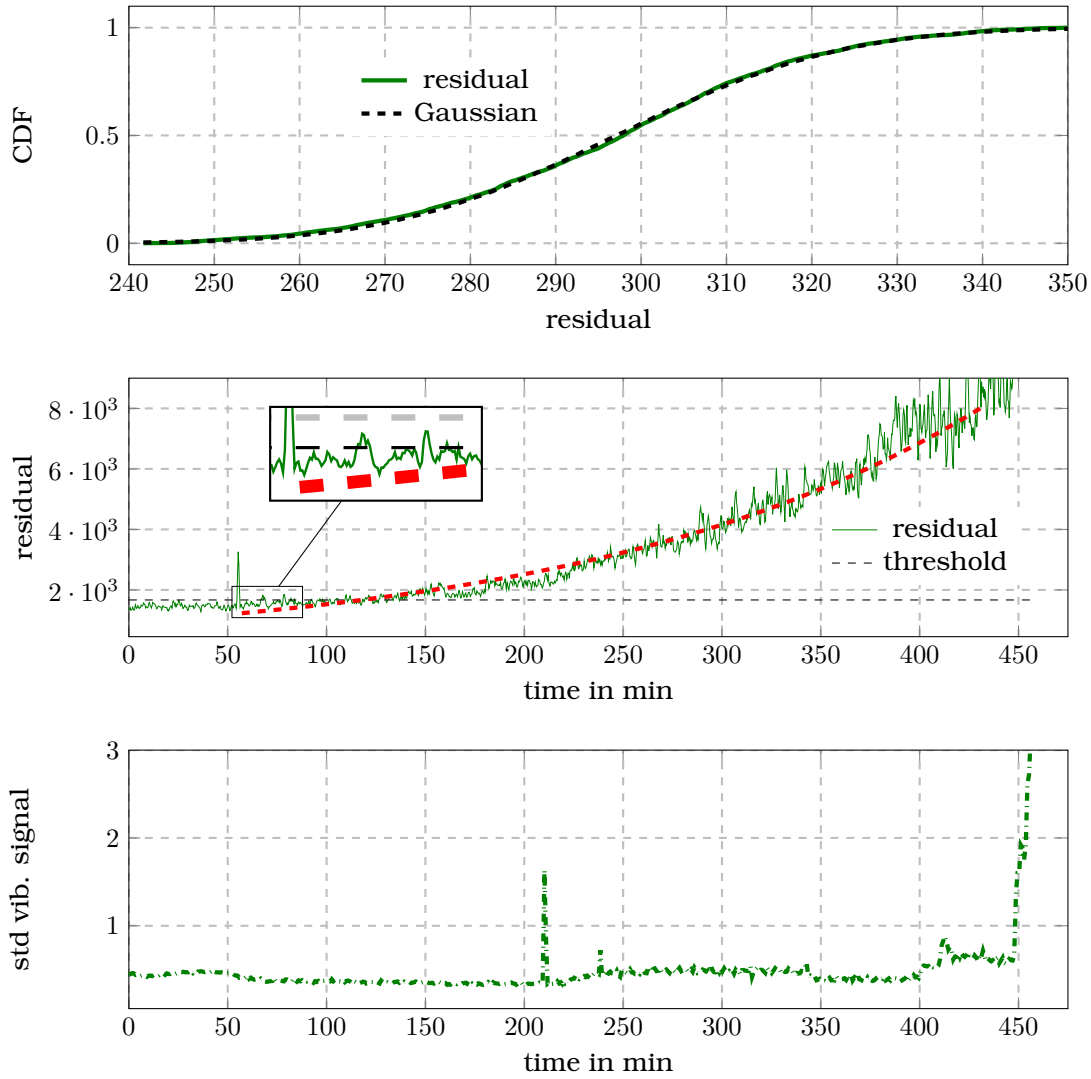


Figure 4.19: Analysis of the data set Bearing1_1 from the PRONOSTIA data set: CDF of the residual for the fault-free case, time evolution of the residual with the fault detection threshold for a run-to failure experiment, and the standard deviation of the vibration signal.

4.2.5 Concluding remarks

A model-based approach for online condition monitoring of rolling element bearings based on efficient and robust numerical differentiation has been proposed. The physical parameters and a fault detection threshold are estimated using fault-free data. The threshold can be designed to guarantee a desired false alarm probability. The detection of sudden degradations and monitoring the evolution of the faults are ensured. The experimental validation using different experimental data sets is promising. These results are very advantageous in terms of implementation simplicity, physical interpretability, low computation complexity, and avoiding training burden.

4.3 Summary and concluding remarks

The previous two sections clearly illustrate that numerical differentiation using orthogonal polynomials (OPs) does not only perform well in simulation studies but also in experimental ones. The successful detection of collisions of a table tennis ball with a mass of $2.7\,\mathrm{g}$ on a magnetically hovering plate with mass equal to $4.7\,\mathrm{kg}$ using the measurement of the vertical displacement of the plate and Jacobi and Laguerre diffs. shows that even in highly disturbed environments higher-order derivatives can be reliably estimated. The approach based on the Lag. diff. does not only achieves a lower detectable minimum falling height of the ball but also considerably reduces the computational burden and the required storage capacities. The implementation of the Jac. diff. for the specially chosen parametrisation shows that the ability to eliminate harmonic disturbances without additional filters is a decisive advantage of these numerical differentiation schemes as previously reported in Kiltz (2017); Kiltz et al. (2013); Kiltz et al. (2014); Kiltz and Rudolph (2013). The use of the Jac. diffs. for the model-based fault detection of rolling element bearings has shown that combining a simple physically motivated model and numerical differentiation can yield powerful tools for fault detection. These tools, validated using different experimental case studies, do not require extensive training with large data sets.

5

On the problem of parameter estimation

Inferring mathematical models from measurements is undoubtedly crucial for most modern control methods. However, the numerical values of the parameters of the models are often not or only poorly known in advance and have to be estimated from measurements. The literature on parameter identification is extensive. For general textbooks on the subject from the control area the reader is referred to Aström and Eykhoff (1971); Garnier and Wang (2008); Isermann (2013a, 2013b); Ljung (1999); Unbehauen and Rao (1987), for instance. Despite the long standing history of the topic, some theoretical questions remain open and new approaches are still developed. Two methods will be discussed and used in this chapter.

An identification approach based on differential-algebraic manipulations in the operational domain has been proposed in Fliess and Sira-Ramírez (2003a). The approach has attracted considerable attention and has been extensively studied and extended in Fliess and Sira-Ramírez (2008); Gehring (2015); Gehring, Rudolph, and Stauch (2012a, 2012b); Gehring, Stauch, and Rudolph (2016); Mboup (2009); Stauch, Gehring, and Rudolph (2013); Ushirobira et al. (2016); Ushirobira, Perruquetti, Mboup, and Fliess (2012, 2013) to name only a few examples. Most of these methods are closely related to the Jacobi differentiator (Jac. diff.) as discussed in Mboup (2009). A detailed list of applications in various engineering fields can be found in Othmane et al. (2022).

Standard gradient estimators (see, e.g., (Slotine & Li, 1991, Sec. 8.7.3) and (Sastry & Bodson, 1989, Sec. 2.5)) for models in the form of linear regression equations (LRE) require the regressor to be persistently excited to ensure the convergence of the estimates. This condition is hardly verified in practice. Recently, the Dynamic Regressor Extension and Mixing (DREM) approach has been proposed in Aranovskiy et al. (2016b, 2017). This parameter estimation method is applicable for models in the form of LRE and consists of two steps. First, new equations are generated using a dynamic operator applied to the available data. Then, these new linearly independent equations are "mixed" to obtain a new regression model on which standard parameter estimation techniques can be applied. The convergence of the approach is ensured without the persistency of excitation condition on the regressor. The approach has attracted a lot of attention (see, e.g., Belov et al. (2018); Ortega et al. (2021, 2019, 2020, 2018); Yi and Ortega (2022)) and has shown to be useful in numerous applications (see, e.g., Aranovskiy et al. (2016a); Bazylev et al. (2017); A. Bobtsov et al. (2017); A. A. Bobtsov et al. (2018); Yi et al. (2018)).

In this chapter, the DREM approach is revisited, and two modifications are proposed. The original approach from Aranovskiy et al. (2016b, 2017) is called standard DREM in the

sequel. The first modification yields less restrictive convergence conditions. In the second modification, a QR decomposition with column pivoting is used to ensure that any excited parameter is adapted even if a subset of the parameters is not. Simple examples are provided to illustrate the modifications and their advantages. A specific example is provided where neither the standard DREM approach nor the modified version yield convergence of the estimation error to zero. Contrarily, an approach motivated by the ideas in Mboup (2009) yields estimates in finite time.

The modified DREM approach is then validated using two different experimental case studies. Therein it is shown that the orthogonal differentiators (orth. diffs.) discussed in Chapters 2 and 3 can be used as operators to generate new linearly independent equations. In the first example, the frequencies of a linear combination of harmonic functions are estimated from a disturbed laboratory measurement. The sampling frequency of the system is $100\,\mathrm{Hz}$, and the highest frequency of interest in the measurement is $80\,\mathrm{Hz}$. The tuning of the differentiators (diffs.) is discussed in detail since it is a challenging problem for this case. The estimation results of this approach are compared to those achieved with a method proposed in Mboup (2009) and implemented using finite-impulse response (FIR) filters. In the second case study, the parameters of a linear time-invariant (LTI) electro-mechanical system are identified from disturbed laboratory measurements. Comparing simulation results with a validation trajectory confirms the excellent results that can be achieved using this approach.

The chapter ends with concluding remarks on the proposed approach and a brief discussion on the simultaneous state and parameter estimation problem.

Chapter content

5.1 Background material on Dynamic Regressor Extension and Mixing 117
5.1.1 Gradient-descent estimator and persistent excitation 118
5.1.2 Dynamic Regressor Extension and Mixing 118
5.2 Two modifications of the DREM approach 121
5.2.1 A first modification of the DREM approach 122
5.2.2 A second modification of the DREM approach 124
5.3 Validation of the approaches 126
5.3.1 Simulation study 127
5.3.2 Experimental case study 1 128
5.3.3 Experimental case study 2 132
5.4 Concluding remarks 135

5.1 Background material on Dynamic Regressor Extension and Mixing

Consider a piecewise constant unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^n$ satisfying the LRE

$$y(t) = \boldsymbol{\phi}^T(t)\boldsymbol{\theta} \tag{5.1}$$

with the known and bounded signals $y : t \mapsto y(t) \in \mathbb{R}$ and $\boldsymbol{\phi} : t \mapsto \boldsymbol{\phi}(t) \in \mathbb{R}^n$. LRE of the form (5.1) appear in numerous problems of parameter estimation and adaptive control (see,

e.g., Åström and Wittenmark (2013); Ljung (1999); Sastry and Bodson (1989)). The online estimation of $\boldsymbol{\theta}$ using the known signals y and $\boldsymbol{\phi}$ is analysed in the following.

For $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\| = \boldsymbol{x}^T\boldsymbol{x}$. The set of nonnegative real numbers is denoted by $\mathbb{R}_{\geqslant 0}$. The set $\mathcal{L}_p$ is defined as

$$\mathcal{L}_p = \left\{ \boldsymbol{f} : \mathbb{R}_{\geqslant 0} \to \mathbb{R}^n \,\middle|\, \int_0^\infty \|\boldsymbol{f}(\tau)\|^p \,\mathrm{d}\tau < \infty \right\},$$

for $p \in (0, \infty)$. The action of an operator $\mathcal{H} : \mathcal{L}_p \to \mathcal{L}_p$ on a signal $\boldsymbol{u} : t \mapsto \boldsymbol{u}(t)$ is denoted by $\mathcal{H}[u](t)$. The identity matrix of size n is denoted by I_n. The adjugate matrix of an $n \times n$ matrix M is denoted by $\mathrm{adj}\{M\}$. A matrix inequality of the form $A_1 > A_2$, where A_1 and A_2 are square matrices of the same dimension, means that $A_1 - A_2 > 0$, i.e., that the matrix $A_1 - A_2$ is positive definite. Similar notations apply to the concepts of positive semi-definiteness, negativ definiteness, and negative semi-definiteness.

5.1.1 Gradient-descent estimator and persistent excitation

The parameter vector $\boldsymbol{\theta}$ can be identified using a gradient-descent estimator (see, e.g., (Slotine & Li, 1991, Sec. 8.7.3) and (Sastry & Bodson, 1989, Sec. 2.5)) of the form

$$\dot{\hat{\boldsymbol{\theta}}}(t) = \Gamma \boldsymbol{\phi}(t) \left(y(t) - \boldsymbol{\phi}^T(t) \hat{\boldsymbol{\theta}}(t) \right) \tag{5.2}$$

with $\Gamma \in \mathbb{R}^{n \times n}$ a positive definite adaptation gain. The parameter estimate error $\boldsymbol{e}_{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ satisfies the linear differential equation

$$\dot{\boldsymbol{e}}_{\boldsymbol{\theta}}(t) = -\Gamma \boldsymbol{\phi}(t) \boldsymbol{\phi}^T(t) \boldsymbol{e}_{\boldsymbol{\theta}}(t), \tag{5.3}$$

whose convergence properties are briefly summarised in the following. The analysis requires the definition of a persistently exciting signal.

Definition 5.9: Persistence of excitation (Sastry and Bodson (1989))

A vector $\boldsymbol{\phi} : \mathbb{R}_+ \to \mathbb{R}^n$ is persistently excited (*PE*) (denoted by $\boldsymbol{\phi} \in$*PE*), if there exist $\alpha_1, \alpha_2, \delta > 0$ such that

$$\alpha_2 I_n \geqslant \int_t^{t+\delta} \boldsymbol{\phi}(\tau) \boldsymbol{\phi}^T(\tau) \mathrm{d}\tau \geqslant \alpha_1 I_n,$$

for all $t \in \mathbb{R}_{\geqslant 0}$.

As discussed, for instance, in (Slotine & Li, 1991, Sec. 8.7.3), the error $\boldsymbol{e}_{\boldsymbol{\theta}}$ exponentially converges to zero, if and only if $\boldsymbol{\phi} \in$ *PE*. Moreover, its norm $\|\boldsymbol{e}_{\boldsymbol{\theta}}\|$ is monotonically non-increasing, i.e.,

$$\|\boldsymbol{e}_{\boldsymbol{\theta}}(t_1)\| \geqslant \|\boldsymbol{e}_{\boldsymbol{\theta}}(t_2)\|, \quad \forall t_1 \leqslant t_2 \in \mathbb{R}_{\geqslant 0}. \tag{5.4}$$

Persistence of excitation is a restrictive condition for most practical applications. In Aranovskiy et al. (2017) an approach called DREM with weaker convergence conditions has been proposed and is now briefly recalled.

5.1.2 Dynamic Regressor Extension and Mixing

The main idea behind the DREM approach is the generation of n new linearly independent one-dimensional LRE and a nonlinear manipulation of the data to independently estimate each

parameter. The first step is achieved using a linear, single-input n-output, bounded-input bounded-output (BIBO)-stable operator $\mathcal{H}$ (see, e.g., (Khalil, 2002, Ch. 5) for more details on input-output stability).

Using the linearity of this operator and applying it to (5.1) yields

$$\boldsymbol{Y}(t) = \Phi(t)\boldsymbol{\theta}, \tag{5.5}$$

with $\boldsymbol{Y}(t) = \mathcal{H}[y](t) \in \mathbb{R}^n$ and $\Phi(t) = \mathcal{H}[\boldsymbol{\phi}^T](t) \in \mathbb{R}^{n\times n}$. Different operators of this kind have been discussed in the DREM literature and are briefly reviewed and commented.

Remark 5.1: Bibliographical comments

Extending the system (5.1) to obtain a square linear system of form (5.5) or more generally a non square overdetermined one is not a novelty of the DREM methodology. Initial ideas can be traced back to early parameter estimation works based on modulating functions in Jouffroy and Reger (2015); Loeb and Cahen (1965); Pearson and Lee (1985); Preisig and Rippin (1993a, 1993b); Shinbrot (1954, 1957), for example. More recently, differential algebraic manipulations in the operational domain have been used to generate new linearly independent equations as shown in Fliess and Sira-Ramírez (2003a, 2008); Gehring (2015); Gehring et al. (2012a, 2012b, 2016); Mboup (2009); Stauch et al. (2013); Ushirobira et al. (2016, 2012, 2013), for example. An advantage of this algebraic approach is the straightforward annihilation of structured perturbations. This can be done in the same manner as for the annihilation of unwanted derivatives in the derivation of diffs. discussed in Section 2.3.2. A connection of these methods of algebraic flavour with orthogonal polynomials (OPs)[a] has been discussed in Mboup et al. (2009); Ushirobira (2018); Ushirobira and Quadrat (2016) and extensively used in Lomakin and Deutscher (2020a, 2020b, 2021, 2022); Othmane et al. (2022, 2020), for instance. In P. Li (2019); Li, Boem, Pin, and Parisini (2020); Pin, Assalone, Lovera, and Parisini (2016); Pin, Lovera, Assalone, and Parisini (2013) a very similar approach to the modulating function method has been discussed. Therein, known signals are convolved with special kernels such that the resulting signals are outputs of stable finite-dimensional state realisations for the integral operators.

A notable difference between the DREM approach and the previously cited works is the estimation of the parameters. For the latter, a well conditioned square or overdetermined linear system of the form (5.5) is sought and solved for the parameter vector $\boldsymbol{\theta}$. In contrast, for the DREM approach a dynamic system is formulated such that the estimation error converges (in finite or infinite time) to zero under some conditions on the matrix Φ. The idea is close to the developments in Kreisselmeier (1977); Lion (1967) for adaptive state and parameter estimation, where new linearly independent regression equations have been obtained using LTI filters and the application of a gradient-descent estimator. These approaches have also been discussed in Jenkins, Annaswamy, and Kojic (2015) and (Narendra & Annaswamy, 2012, Sec. 6.5), for instance. As for the simple estimator in (5.3) the latter works require the regressor to be *PE*. The only difference is that the convergence rate can be made arbitrarily fast by increasing the adaptation gain and carefully choosing the operators for the generation of the new equations. However, this remains questionable since it may yield numerical instabilities and a scheme that is less

robust with respect to unmodeled dynamics (see, e.g., Jenkins et al. (2015); Ortega et al. (2020)). The DREM approach in contrast is a method with relaxed convergence conditions and an improved transient performance as will be discussed.

[a]See the survey Othmane et al. (2022) for a detailed discussion on the historical background of Jac. diffs. and an extensive list of applications in the context of parameter estimation.

DREM-operators

Different formulation for the BIBO-stable operators $\mathcal{H}$ have been introduced in the literature. In Aranovskiy et al. (2017), exponentially stable low-pass filters of first order and pure delay operators are discussed. In Ortega et al. (2021), a generalisation of the latter two operators is proposed in form of linear time-varying operators

$$\begin{aligned}\dot{\boldsymbol{x}}_i(t) &= A_i(t)\boldsymbol{x}_i(t) + \boldsymbol{b}_i(t)u(t)\\ z_i(t) &= \boldsymbol{c}_i^T(t)\boldsymbol{x}_i(t) + d_i(t)u(t) + \mu_i(t)u(t-T_i),\end{aligned}$$

with $A_i : t \mapsto A_i(t) \in \mathbb{R}^{n_i\times n_i}$, $n_i \in \mathbb{N}$, $b_i : t \mapsto b_i(t) \in \mathbb{R}^{n_i}$, $c_i : t \mapsto c_i(t) \in \mathbb{R}^{n_i}$, $d_i : t \mapsto d_i(t) \in \mathbb{R}^{n_i}$, $\mu_i : t \mapsto \mu_i(t) \in \mathbb{R}$ time-varying functions, $T_i \geqslant 0$, and $i \in \{1,\dots,n-1\}$.

In (Ortega et al., 2021, Prop. 4), it is shown that Kreisselmeier's regressor extension developed in Kreisselmeier (1977) is a special case of the latter operators. Linear time-varying operators are used in Ortega et al. (2018) to attach a Luenberger observer interpretation to this parameter estimation approach.

Parameter estimation

Multiplying both sides of (5.5) with the adjugate of $\Phi(t)$ yields[1]

$$\boldsymbol{\mathcal{Y}}(t) = \Delta(t)\boldsymbol{\theta} \tag{5.6a}$$

with

$$\boldsymbol{\mathcal{Y}}(t) = \operatorname{adj}\{\Phi(t)\}\,\boldsymbol{Y}(t) \quad \text{and} \quad \Delta(t) = \det\{\Phi(t)\}. \tag{5.6b}$$

The relation in (5.6a) can also be written componentwise for each element of the parameter vector $\boldsymbol{\theta}$ as $\mathcal{Y}_i(t) = \Delta(t)\theta_i$, $i \in \{1,\dots,n\}$. The parameters θ_i, $i \in \{1,\dots,n\}$, can be estimated in continuous-time by the gradient-descent estimator

$$\dot{\hat{\theta}}_i(t) = \gamma_i\Delta(t)\left(\mathcal{Y}_i(t) - \Delta(t)\hat{\theta}_i(t)\right), \quad \gamma_i \in \mathbb{R}_{>0}, \tag{5.7}$$

proposed in Aranovskiy et al. (2017). The estimation errors $e_{\boldsymbol{\theta},i} = \theta_i - \hat{\theta}_i$, $i \in \{1,\dots,n\}$, satisfy the differential equations $\dot{e}_{\boldsymbol{\theta},i}(t) = -\gamma_i\Delta^2(t)e_{\boldsymbol{\theta},i}(t)$. The individual parameter errors are monotonically non-increasing, i.e.,

$$|e_{\boldsymbol{\theta},i}(t_1)| \geqslant |e_{\boldsymbol{\theta},i}(t_2)|, \quad \forall t_1 \leqslant t_2 \in \mathbb{R}_{\geqslant 0}, \tag{5.8}$$

and $\lim_{t\to\infty} e_{\boldsymbol{\theta},i}(t) = 0$ if and only if $\Delta \notin \mathcal{L}_2$. The convergence of each error $e_{\boldsymbol{\theta},i}$ can be adjusted by varying γ_i. If $\Delta \in PE$ the convergence is exponential.

According to Aranovskiy et al. (2017), this estimation approach has three main advantages over the standard gradient-descent estimator. First, the individual parameter errors in (5.8)

[1]Recall that for a square $n \times n$ matrix M it holds that $\operatorname{adj}\{M(t)\}\,M(t) = \det\{M(t)\}I_n$.

are monotonically non-increasing, a property that is strictly stronger than the monotonicity of their norm as in the standard gradient-descent estimator in (5.4). Second, the parameter convergence is established without the restrictive *PE* assumption. A less restrictive condition, the non-square integrability assumption, is required. Third, the convergence rates of DREM can be adapted for each single parameter by adapting γ_i.

Remark 5.2: Use of orth. diffs. in DREM

It can be recognised that the orth. diffs. derived in Chapter 2 can be used as operators to get a system of the form (5.5) and the gradient-descent estimator in (5.7) can then be used to estimate the parameters. The detailed discussions concerning tuning guidelines and implementation issues in Chapter 3 can be used to efficiently parametrise the operators. This will be discussed in detail on the basis of experimental examples in the next section. Beforehand, an extension of this approach is proposed for general BIBO operators and generalises the method discussed in Othmane et al. (2020).

5.2 Two modifications of the DREM approach

In this section, a generalisation of the gradient estimator recalled in Section 5.1.2 in form of a linear time-varying observer is proposed. Consider the linear time-varying system

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{0}, \quad \boldsymbol{Y}(t) = \Phi(t)\boldsymbol{z}(t) \tag{5.9}$$

with the state $\boldsymbol{z}(t) \in \mathbb{R}^n$ modelling the sought parameters $\boldsymbol{\theta}$, the bounded output $\boldsymbol{Y}(t) \in \mathbb{R}^n$ satisfying (5.5), and the initial condition $\boldsymbol{z}(t_0) = \boldsymbol{\theta}$.

The parameter vector $\boldsymbol{z}$ shall be estimated using an observer

$$\dot{\hat{\boldsymbol{z}}}(t) = L(t)\left(\boldsymbol{Y}(t) - \Phi(t)\hat{\boldsymbol{z}}(t)\right), \tag{5.10}$$

with state $\hat{\boldsymbol{z}}(t)$, initial condition $\hat{\boldsymbol{z}}(t_0) = \hat{\boldsymbol{z}}_0 \in \mathbb{R}^n$, and a time-varying gain $L : t \mapsto L(t)$. The gain shall be computed such that the estimation error $\boldsymbol{e} = \boldsymbol{z} - \hat{\boldsymbol{z}}$ satisfying the differential equation

$$\dot{\boldsymbol{e}}(t) = -L(t)\left(\boldsymbol{Y}(t) - \Phi(t)\hat{\boldsymbol{z}}(t)\right) = -L(t)\Phi(t)\left(\boldsymbol{z}(t) - \hat{\boldsymbol{z}}(t)\right) = -L(t)\Phi(t)\boldsymbol{e}(t) \tag{5.11}$$

converges to zero.

Two approaches for the computation of L are discussed. First, it will be shown that the convergence condition $\Delta \notin \mathcal{L}_2$ for the estimator (5.7) can be generalised to $\Delta \notin \mathcal{L}_p$ for an arbitrary $p \in (0, \infty)$. However, similarly to the initially discussed approach, this extension suffers from the following problem: A poor excitation of some parameters in the system during a time interval $\mathcal{I}_h(t) = [t-h, t]$, $h > 0$, yields a rank deficient matrix $\Phi(t)$, i.e., $\det\{\Phi(\tau)\} \approx 0$, $\tau \in \mathcal{I}_h(t)$. Hence, parameters will not be adapted during this interval. An approach partially overcoming this problem is then proposed. The key idea is to decompose the parameter vector into excited parameters and non-excited ones, at every time instant t. Then, the excited parameters can always be adapted. This can be achieved using a QR factorisation with column pivoting. The resulting approach is called generalised DREM (gDREM) in the sequel.

5.2.1 A first modification of the DREM approach

The first idea for a modification of DREM is based on the generalisation of (5.6) to

$$L(t)\Phi(t) = f_p(\Phi(t))\Gamma, \tag{5.12a}$$

with $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$ a diagonal matrix with positive diagonal entries and the function

$$f_p(\Phi(t)) = |\det\{\Phi(t)\}|^p, \quad p \in (0, \infty). \tag{5.12b}$$

A matrix L that satisfies (5.12) is[2]

$$L(t) = \begin{cases} f_p(\Phi(t))\Gamma\Phi^{-1}(t), & \text{for} \quad \det\{\Phi(t)\} \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.13}$$

The estimation error $\boldsymbol{e}(t) = \boldsymbol{z}(t) - \hat{\boldsymbol{z}}(t)$ satisfies the differential equation (5.11), which for this choice of L becomes $\dot{\boldsymbol{e}}(t) = -f_p(\Phi(t))\Gamma\boldsymbol{e}(t)$. Its solution is

$$\boldsymbol{e}(t) = \exp\left(-\Gamma \int_{t_0}^{t} f_p(\Phi(\tau))\, \mathrm{d}\tau\right)(\boldsymbol{z}(t_0) - \hat{\boldsymbol{z}}(t_0)).$$

The individual components e_i, $i \in \{1, \ldots, n\}$, of the error $\boldsymbol{e}$ can be written as

$$e_i(t) = \exp\left(-\gamma_i \int_{t_0}^{t} f_p(\Phi(\tau))\, \mathrm{d}\tau\right)(z_i(t_0) - \hat{z}_i(t_0)).$$

The derivations above establish the following result and it is easy to verify that for $p = 2$ the DREM approach with the adjugate matrix is recovered.

Proposition 5.1: First modification of DREM

Consider the estimation error $\boldsymbol{e} = \boldsymbol{z} - \hat{\boldsymbol{z}}(t)$ with $\boldsymbol{z}$ from (5.9) and $\hat{\boldsymbol{z}}(t)$ from (5.10), where the gain L is computed according to (5.13). The following equivalence holds

$$\lim_{t\to\infty} e_i(t) = 0 \Leftrightarrow t \mapsto \det\{\Phi(t)\} \notin \mathcal{L}_p, \quad p \in (0, \infty),$$

for all components e_i, $i \in \{1, \ldots, n\}$, of the error $\boldsymbol{e}$. The convergence is exponential if $t \mapsto \det\{\Phi(t)\} \in$ *PE*.

The condition $t \mapsto \det\{\Phi(t)\} \notin \mathcal{L}_p$, $p \in (0, \infty)$, is less restrictive than the condition recalled in Section 5.1.2 requiring that $t \mapsto \det\{\Phi(t)\}$ is not square integrable, as demonstrated in the following example for $p = 1$.

Example 5.9 Example for first DREM modification

Consider the regressor

$$\boldsymbol{\phi}(t) = \left[\frac{-(t+1)\cos(t+1) + \cos(1) - \sin(1) + \sin(t+1)}{t+1}, \frac{1}{t+1}\right]^T$$

the components of which are bounded for all $t \geqslant 0$. Moreover, the second component

[2] It can be recognised, that the gain computed using (5.13) corresponds to a solution of the equation $L(t)\Phi(t) = f(\Phi(t))\,\Gamma$. The QR or the Singular Value decompositions can be used to compute the inverse of $\Phi(t)$ as discussed, for instance, in Golub and Van Loan (2013) or Press, Teukolsky, Vetterling, and Flannery (2007) in the context of linear algebraic equations.

converges to zero for $t \to \infty$. For the application of the previously discussed approaches a system of the form (5.5) is required, which can be achieved by defining

$$\Phi(t) = \begin{bmatrix} \phi(t) \\ \dot{\phi}(t) \end{bmatrix},$$

i.e., the system is augmented by the derivative of the regressor[a]. This is a reasonable choice since the orth. diffs. discussed in Chapters 2 and 3 can be used to approximate these signals. The determinant $\det\{\Phi(t)\}$ of $\Phi(t)$ satisfies $\det\{\Phi(t)\} = \operatorname{sinc}(t)$ for all t, where sinc is the sinc function summarised in #A.19. The function sinc satisfies $\operatorname{sinc} \in \mathcal{L}_2$ and $\operatorname{sinc} \notin \mathcal{L}_1$. Thus, the estimation error of the standard DREM approach would fail to converge to zero[b]. Conversely, the modified approach with $p = 1$ ensures a convergence to zero.

[a] Observe that the augmentation of general LRE by derivatives of the regressors yields conditions on the Wronskian of the regressor for the error to converge to zero.

[b] By increasing the adaptation gain Γ the estimation error of the standard DREM approach can be made arbitrarily small. However, this can yield numerical instabilities as for the approach in Jenkins et al. (2015) using the methods from Kreisselmeier (1977); Lion (1967).

Moreover, a bounded function f satisfies

$$f^2(t) \leqslant |f(t)| \sup_{\tau \in \mathbb{R}_{\geqslant 0}} |f(\tau)|$$

for all $t \geqslant 0$. Thus, every function $f \in \mathcal{L}_1$ is also in $\mathcal{L}_2$. Hence, if the determinant of Φ is $\Delta \in \mathcal{L}_1$ neither of the approaches converge. This is again illustrated using an example.

Example 5.10 Counter example for first DREM modification

Consider the estimation of the parameters $d, \omega \in \mathbb{R}_{>0}$ of the signal $y(t) = \mathrm{e}^{-dt} \sin \omega t$. As discussed in Mboup (2009) the signal satisfies the differential equation

$$\ddot{y} + 2\theta_1 \dot{y} + \theta_2 y = 0, \quad \theta_1 = d, \quad \theta_2 = \omega^2 + d^2.$$

According to (Mboup, 2009, Sec. 3) the parameters θ_1 and θ_2 are linearly identifiable[a]. On the other hand, d and ω, which are algebraic functions of θ_1 and θ_2, are weakly linearly identifiable. By defining the regressor $\phi(t) = [2\dot{y}, y]^T$ an LRE of the form (5.1) is obtained[b]. Systems of the form (5.5) can be obtained using a delay or a differentiation operator. Denote by Φ_1 and Φ_2 the matrices of systems of the form (5.5) obtained using the derivative of ϕ and a delayed version of ϕ with a delay $h > 0$, respectively. Their respective determinants Δ_1 and Δ_2 satisfy

$$\lim_{t\to\infty} \int_0^t \Delta_1^2(\tau)\mathrm{d}\tau = \frac{\omega^4}{d} \qquad \lim_{t\to\infty} \int_0^t |\Delta_1(\tau)|\mathrm{d}\tau = \frac{\omega^2}{d},$$
$$\lim_{t\to\infty} \int_0^t \Delta_2^2(\tau)\mathrm{d}\tau = \frac{\omega^2 \sin^2(h\omega)\mathrm{e}^{2dh}}{d}, \qquad \lim_{t\to\infty} \int_0^t |\Delta_2(\tau)|\mathrm{d}\tau = \frac{\omega|\sin(h\omega)|\mathrm{e}^{dh}}{d}.$$

Thus, none of these approaches yields a convergence of the estimation error to zero. Contrarily, the algebraic approach from Mboup (2009) yields very good estimates.

[a] The reader is referred to Fliess, Mboup, Mounier, and Sira-Ramírez (2003); Fliess and Sira-Ramírez (2003a) for an in depth discussion of the identifiability concepts used in this example.

[b]Using derivatives of y in the regressor does not present any issue since orth. diffs. can be used to filter the entire equation and computable signals are then obtained.

While this modification of DREM yields weaker convergence conditions, no adaptation is performed as soon as one parameter is not excited. This shall be overcome in the next modification of DREM, which is an extension of the first one.

5.2.2 A second modification of the DREM approach

Suppose $\Phi(t)$ has rank[3] $\kappa(t) \in \mathbb{N}$. Using a QR factorisation with column pivoting (see, e.g., (Golub & Van Loan, 2013, Sec. 5.2)), the matrix $\Phi(t)$ can be decomposed as

$$\Phi(t)\Pi(t) = Q(t)R(t) = Q(t)\begin{bmatrix} R_1(t) & R_2(t) \\ 0_{\bar{\kappa}(t)\times\kappa(t)} & 0_{\bar{\kappa}(t)\times\kappa(t)} \end{bmatrix}, \quad \bar{\kappa}(t) = n - \kappa(t) \tag{5.14}$$

at each instant t, with $Q(t) \in \mathbb{R}^{n\times n}$ an orthogonal matrix, $R(t) \in \mathbb{R}^{n\times n}$ an upper triangular matrix. The matrices $R_1(t)$ and $R_2(t)$ of dimensions $\kappa(t) \times \kappa(t)$ and $\kappa(t) \times \bar{\kappa}(t)$, respectively, contain the information stemming from the excited parameters. These can be extracted from the parameter vector using the permutation matrix $\Pi(t)$. If $\Phi(t)$ has full rank, i.e., $\kappa(t) = n$, then $R(t) = R_1(t)$.

As opposed to the standard DREM with the adjugate operator in (5.6), or the generalisation proposed in (5.12), the extension and mixing is now performed in a different but closely related manner.

Proposition 5.2: Second modification of DREM

Consider the estimation error $\boldsymbol{e} = \boldsymbol{z} - \hat{\boldsymbol{z}}$ with $\boldsymbol{z}$ from (5.9) and $\hat{\boldsymbol{z}}$ from (5.10) where the gain $t \mapsto L(t)$ is computed at each instant t as

$$L(t) = -|\det\{R_1(t)\}|^p \Gamma K(t), \quad p \in (0, \infty), \tag{5.15a}$$

with $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$ a diagonal matrix with positive entries and

$$\begin{aligned} K(t) &= \begin{cases} \Pi(t)G(t)Q^T(t), & \text{for} \quad \big(\kappa(t) > 0 \wedge R_2(t) = 0_{\kappa(t)\times\bar{\kappa}(t)}\big) \vee \kappa(t) = n, \\ 0_{n\times n}, & \text{otherwise}, \end{cases} \\ G(t) &= \begin{bmatrix} R_1^{-1}(t) & 0_{\kappa(t)\times\bar{\kappa}(t)} \\ 0_{\bar{\kappa}(t)\times\kappa(t)} & 0_{\bar{\kappa}(t)\times\kappa(t)} \end{bmatrix}, \end{aligned} \tag{5.15b}$$

where $\Pi(t)$, $R(t)$, $Q(t)$, $R_1(t)$, and $R_2(t)$ are given in (5.14). The errors e_i, $i \in \{1, \ldots, n\}$ satisfy decoupled differential equations. If

$$\det\{\Phi(t)\} \notin \mathcal{L}_p \tag{5.16}$$

then $\lim_{t\to\infty} e_i(t) = 0$. Moreover, the convergence is exponential if $\det\{\Phi(t)\} \in PE$.

Proof. For notational simplicity, let

$$r_1(t) = |\det\{R_1(t)\}|^p .$$

[3]For the determination of the rank of a matrix using a QR factorisation the reader is referred to (Golub & Van Loan, 2013, Sec. 5.4.3), for example.

The estimation error $\boldsymbol{e}(t) = \boldsymbol{z}(t) - \hat{\boldsymbol{z}}(t)$ satisfies the differential equation (5.11) which for L computed as in (5.15) becomes

$$\begin{aligned}
\dot{\boldsymbol{e}}(t) &= r_1(t)L(t)\Phi(t)\boldsymbol{e}(t) \\
&= -r_1(t)\Gamma \begin{cases} \Pi(t)G(t)Q^T(t)Q(t)R(t)\Pi^T(t), & \text{for} \quad \big(\kappa(t) > 0 \wedge R_2(t) = 0_{\kappa(t)\times\bar{\kappa}(t)}\big) \vee \kappa(t) = n, \\ 0, & \text{otherwise}, \end{cases} \\
&= -r_1(t)\Gamma \begin{cases} \Pi(t)G(t)R(t)\Pi^T(t), & \text{for} \quad \big(\kappa(t) > 0 \wedge R_2(t) = 0_{\kappa(t)\times\bar{\kappa}(t)}\big) \vee \kappa(t) = n, \\ 0, & \text{otherwise}, \end{cases} \\
&= -r_1(t)\Gamma \begin{cases} \Pi(t)J(t)\Pi^T(t), & \text{for} \quad \big(\kappa(t) > 0 \wedge R_2(t) = 0_{\kappa(t)\times\bar{\kappa}(t)}\big) \vee \kappa(t) = n, \\ 0, & \text{otherwise}, \end{cases}
\end{aligned}$$

with

$$J(t) = \begin{bmatrix} I_{\kappa(t)\times\kappa(t)} & 0_{\kappa(t)\times\bar{\kappa}(t)} \\ 0_{\bar{\kappa}(t)\times\kappa(t)} & 0_{\bar{\kappa}(t)\times\kappa(t)} \end{bmatrix}.$$

Since $\Pi(t)$ is a permutation matrix, $\Pi(t)J(t)\Pi^T(t)$ is a diagonal matrix with elements equal to zero or one for all t. Thus, the errors e_i, $i \in \{1, \ldots, n\}$, satisfy decoupled differential equations

$$\dot{e}_i(t) = -\gamma_i r_1(t) f_i(t) e_i(t)$$

with $f_i(t) \in \{0, 1\}$. Recall that for $\det\{\Phi(t)\} \neq 0$ the rank of $\Phi(t)$ is n and $R_1 = R(t)$. Furthermore, if $r_1(t) = 0$ then $\det\{\Phi(t)\} = 0$. Thus, by observing that

$$0 \leqslant |\det\{\Phi(t)\}| \leqslant r_1(t) f_i(t)$$

it can be concluded that if $|\det\{\Phi(t)\}| \notin \mathcal{L}_p$, then $\lim_{t\to\infty} \int_0^t r_1(\tau) f_i(\tau) \mathrm{d}\tau = \infty$ and

$$\lim_{t\to\infty} e_i(t) = \exp\left(-\gamma_i \int_{t_0}^{t} f_i(\tau) r_1(\tau)\right)(z_i(t_0) - \hat{z}_i(t_0)) = 0,$$

which concludes the proof. The exponential convergence claim follows from the definition of *PE*. □

Computing the observer gain L at each time instant as in (5.15) corresponds to an extension of the result in Proposition 5.1. This can be seen by observing that when $\Phi(t)$ is regular, both gains are equal. They differ only for rank deficient $\Phi(t)$ with $R_2(t) = 0_{\kappa(t)\times\bar{\kappa}(t)}$. While finding conditions on $\Phi(t)$ such that $R_2(t) = 0_{\kappa(t)\times\bar{\kappa}(t)}$ remains an open problem, the considered examples in the following show that this approach presents numerous advantages. The following example shows that the condition (5.16) is only a necessary condition for the convergence to zero of all estimation errors.

Example 5.11 Second DREM modification

Consider the regressor

$$\phi(t) = \begin{bmatrix} \frac{1}{t+1} & r(t) \end{bmatrix}, \quad r(t) = \begin{cases} 1, & \text{for} \quad t \leqslant \frac{1}{2}, \\ 0, & \text{otherwise}. \end{cases}$$

The matrix Φ can be generated using a delay $h \in (0, 1)$, for instance, to get

$$\Phi(t) = \begin{bmatrix} \frac{1}{t+1} & r(t) \\ \frac{1}{t+1-h} & r(t-h) \end{bmatrix}.$$

It is straightforward to see that $\det\{\Phi(t)\}$ vanishes in finite time. Thus, the standard DREM estimation approach would fail to converge to zero. However, the QR decomposition as discussed above yields the following error dynamics for the two parameters

$$\dot{e}_1(t) = \begin{cases} -\gamma_1 \left| \frac{1}{(t+1-h)} \right|^p e_1(t), & \text{for} \quad t \in [0, h], \\ -\gamma_1 \left| \frac{1}{(t+1)(t+1-h)} \right|^p e_1(t), & \text{for} \quad t \in \left(h, \frac{1}{2} + h\right], \\ -\frac{\gamma_1}{t+1} e_1(t), & \text{otherwise}, \end{cases}$$

$$\dot{e}_2(t) = \begin{cases} -\gamma_2 \left| \frac{1}{t+1-h} \right|^p e_2(t), & \text{for} \quad t \in [0, h], \\ -\gamma_2 \left| \frac{1}{(t+1)(t+1-h)} \right|^p e_2(t), & \text{for} \quad t \in \left(h, \frac{1}{2} + h\right], \\ 0, & \text{otherwise}. \end{cases}$$

Thus, the error e_1 converges to zero while the error e_2 remains constant for $t > 1$.

The standard DREM approach recalled in Section 5.1.2 requires the adjugate matrix of $\Phi(t)$, which is the transpose of the matrix of the co-factors of $\Phi(t)$. The computation of the adjugate from this definition requires the calculation of n^2 determinants of order $n - 1$. This algorithm has a complexity of $O(n^4)$. A more efficient and numerically stable algorithm is the computation of the adjugate using a QR decomposition as discussed in Stewart (1998). Thus, the slight modification of the gain in Proposition 5.2 for the parameter estimation does not noticeably increase the complexity of the required computations. Hence, when the parameter p in (5.12b) is chosen equal to 2, this modification of the DREM approach does not yield any disadvantages compared to the initial formulation in Section 5.1.2.

Remark 5.1: Nonlinearly parametrised regressions

In Aranovskiy et al. (2017) the use of the DREM technique for nonlinearly parametrised regressions with factorisable nonlinearities has been discussed. The approach can also be extended using the observer proposed here.

5.3 Validation of the approaches

In the following, the gDREM approach is validated using three case studies. First, a simulation is provided where only the excited parameters are updated. The non excited ones are kept constant. Then, two experimental cases studies where gDREM based on Jac. diffs. are discussed. In the sequel, all dynamic systems are discretised using the Euler backward method.

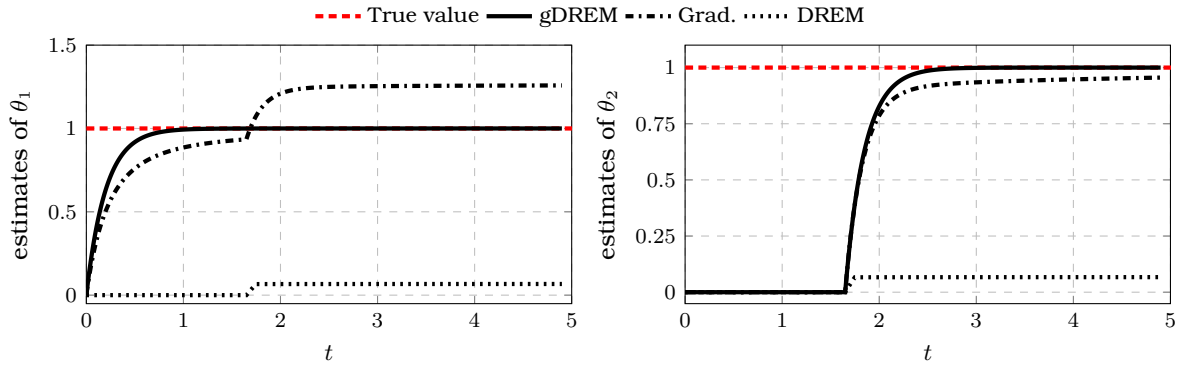


Figure 5.1: Time evolution of the parameter estimates for the system in (5.17) with $\Gamma_{\text{gDREM}} = \Gamma_{\text{DREM}} = \Gamma_{\text{Grad}} = [5, 5]$ using three different approaches.

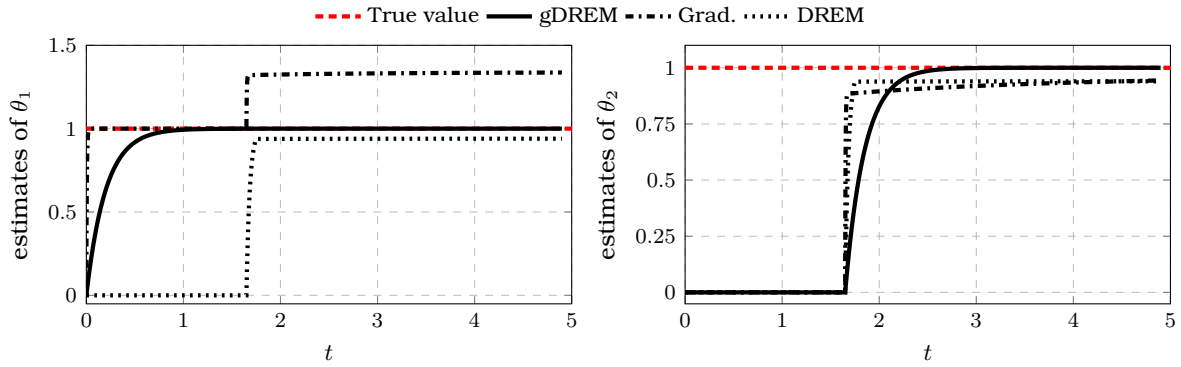


Figure 5.2: Time evolution of the parameter estimates for the system in (5.17) with $\Gamma_{\text{gDREM}} = [5, 5]$ and $\Gamma_{\text{DREM}} = \Gamma_{\text{Grad}} = [200, 200]$ using three different approaches.

5.3.1 Simulation study

Consider the system

$$y(t) = \boldsymbol{\phi}^T(t) \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad \boldsymbol{\phi}(t) = \begin{bmatrix} \frac{1}{t+1} & r(t) \end{bmatrix}^T, \quad r(t) = \begin{cases} 0, & \text{for} \quad t \leqslant 1.75, \\ 1, & \text{otherwise.} \end{cases} \tag{5.17}$$

The regressor $\boldsymbol{\phi}$ is not *PE* since $\lim_{t\to\infty} \frac{1}{t+1} = 0$. To apply DREM or gDREM one more equation has to be generated. This is done here using a delay operator with a delay $h = 1/2$. It can be verified that the convergence conditions for the gradient estimator and the standard DREM are not fulfilled. However, the convergence conditions for gDREM with $p = 1$ are satisfied. In the simulation $\theta_1 = \theta_2 = 1$.

The time evolution of the estimates for two different values of Γ is shown in the Fig. 5.1 and Fig. 5.2. It is clear that the gDREM estimates converge to the true values while the remaining two approaches do not. Here it is important to note that the estimate of θ_1 for gDREM starts to converge to its true value from $t = 0$ on. The estimate of θ_2 starts to converge beginning from $t = 1.75$. This once again confirms the importance of the approach using the QR decomposition to detect the excited parameters.

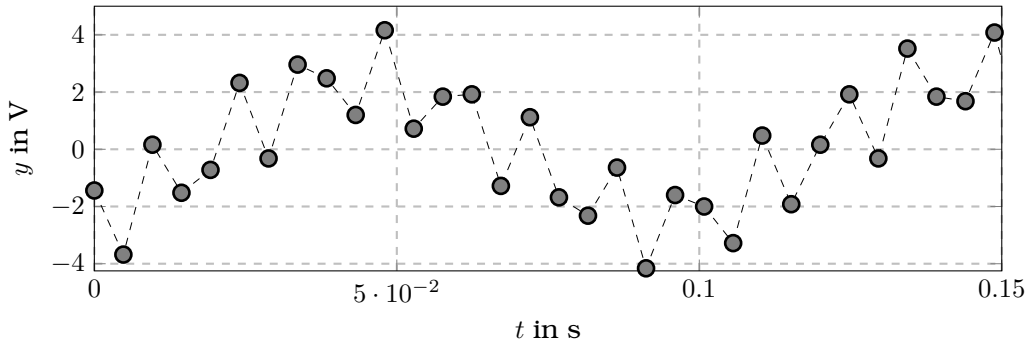


Figure 5.3: Time evolution of the measurement for the experimental case study 1 in the time interval $[0\,\mathrm{s}, 0.15\,\mathrm{s}]$: The discrete measurements are displayed by the markers.

5.3.2 Experimental case study 1

The DREM approach is first validated using a frequency estimation problem. To this end, assume a signal

$$y : t \mapsto y(t) = a_0 \sin(\omega_1 t + \phi_1) + a_1 \sin(\omega_2 t + \phi_2), \quad a_0, a_1, \phi_1, \phi_2 \in \mathbb{R}, \tag{5.18}$$

is known at discrete time instances. The frequencies $\omega_1, \omega_2 > 0$ shall be estimated. The signal y is generated in the laboratory using a signal generator where $\omega_1 = 20\pi\,\mathrm{rad/s}$ and $\omega_2 = 80\pi\,\mathrm{rad/s}$. The signal is recorded using a sampling period of $t_\mathrm{s} = 5\,\mathrm{ms}$. The corresponding Nyquist frequency is $100\,\mathrm{Hz}$. The recorded signal is shown in Fig. 5.3. The low sampling rate in comparison to ω_2 makes the problem challenging especially when diffs. are used.

From a signal processing perspective the Fast Fourier Transform (see, e.g., (Oppenheim & Schafer, 1975, Ch. 6)) can be used for the extraction of the frequencies from logged measurements. However, this can be computationally intensive for an online estimation. Numerous other approaches better suited for online applications have been developed. Algebraic parameter estimation approaches[4] based on differential-algebraic manipulations in the operational domain inspired by the fundamental ideas in Fliess and Sira-Ramírez (2003a) have been applied in Liu et al. (2008); Trapero, Sira-Ramírez, and Batlle (2007); Ushirobira et al. (2012) for the estimation of a single frequency and in Trapero, Sira-Ramírez, and Batlle (2008); Ushirobira et al. (2016, 2013) for the estimation of two frequencies simultaneously with the signal amplitudes and phases. A different approach for the estimation of multiple frequencies relies on adaptive observers that show global or semi-global stability properties. These approaches have been discussed in Chen, Pin, and Parisini (2014); Hou (2012); Marino and Tomei (2002); Xia (2002), for instance. A convergence of the estimates in finite-time can also be achieved using integral operators as proved by Chen, Li, Pin, and Parisini (2016). Therein, the measurement is convolved with a kernel and its derivatives to overcome the problem of unknown derivatives. In this sense, the approach is closely related to the algebraic identification method. In Fedele and Coluccio (2010) modulating functions initially introduced[5] in Shinbrot (1954, 1957) have been used for the recursive estimation of the frequencies. Due to the importance of modulating functions, their definition is recalled in the following.

[4] In Mboup (2009) a least squares interpretation is attached to an algebraic identification approach.
[5] See Preisig and Rippin (1993a, 1993b) for a detailed review on modulating functions.

Definition 5.10: Modulation function (Preisig and Rippin (1993a))

A function $\phi : [0, T] \to \mathbb{R}$ is called a modulating function (of order $\alpha \in \mathbb{N}$) if it is sufficiently smooth and if, for some fixed T

$$\phi^{(i)}(0) = \phi^{(i)}(T) = 0,$$

for all $i \in \{0, \ldots, \alpha - 1\}$.

From Definition 2.8 of Jac. diff. $g^{(\alpha,\beta)}_{N,T,\vartheta}$, it is clear that it is a modulating function of order $\min(\alpha, \beta)$. This property of Jac. diffs. has been used in Liu, Laleg-Kirati, Perruquetti, and Gibaru (2014) for the simultaneous estimation of the state and unknown input of a nonlinear system. It is now used for the estimation of the frequencies ω_1 and ω_2.

It is well known that the signal y from (5.18) satisfies the differential equation

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}^2 + \omega_1^2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t}^2 + \omega_2^2\right) y = 0,$$

or equivalently

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}^4 + \theta_1 \frac{\mathrm{d}}{\mathrm{d}t}^2 + \theta_2\right) y = 0, \quad \theta_1 = \left(\omega_1^2 + \omega_2^2\right), \quad \theta_2 = \omega_1^2 \omega_2^2. \tag{5.19}$$

Thus, according to Mboup (2009) the parameters θ_1 and θ_2 are linearly identifiable. On the other hand, the parameters ω_1 and ω_2, which are algebraic functions of θ_1 and θ_2, are weakly linearly identifiable. The latter equation is used in the following for the estimation of θ_1 and θ_2. Denote by $\hat{\theta}_1$ and $\hat{\theta}_2$ the corresponding estimates. The estimates of the frequencies ω_1 and ω_2, denoted in the sequel by $\hat{\omega}_1$ and $\hat{\omega}_2$, respectively, can then be computed as the positive roots of the polynomial $\lambda^2 + \hat{\theta}_1 \lambda + \hat{\theta}_2$.

Let $g_i = g^{(\alpha_i,\beta_i)}_{0,T_i,\vartheta}$ be a Jac. diff. with $\alpha_i, \beta_i > 3$ and denote by

$$\hat{y}_i^{(n)}(t) = \int_0^T g_i^{(n)}(\tau) y(t - \tau) \mathrm{d}\tau.$$

For notational simplicity, the explicit time dependence of signals is dropped in the sequel. Convolving both sides of the differential equation (5.19) with g_i yields[6],

$$\hat{y}_i^{(4)} + \theta_1 \ddot{\hat{y}}_i + \theta_2 \hat{y} = 0 \quad \Leftrightarrow \quad \hat{y}_i^{(4)} = \boldsymbol{\phi}_i^T \boldsymbol{\theta},$$

with $\boldsymbol{\phi_i} = -[\ddot{\hat{y}}_i, \hat{y}_i]^T$ and $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$.

Two approaches for the estimation of the frequencies are now discussed:

1. Using $N_{\mathrm{eq}} \geqslant 2$ different Jac. diff. a linear system of equations

$$\boldsymbol{Y} = \Phi \boldsymbol{\theta}, \tag{5.20}$$

with $\boldsymbol{Y} = [\hat{y}_1^{(4)}, \ldots, \hat{y}_{N_{\mathrm{eq}}}^{(4)}]^T$ and $\Phi = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{N_{\mathrm{eq}}}]^T$ is generated. For $N_{\mathrm{eq}} > 2$ the system is overdetermined and a least squares solution can be computed. In all cases the solution can be computed using a Singular Value Decomposition as described in Press et al.

[6] This equation can also be recovered using differential-algebraic manipulations in the operational domain as discussed in Mboup (2009). This can be easily seen by recalling the differential-algebraic derivation of Jac. diffs. in Section 2.4.3.

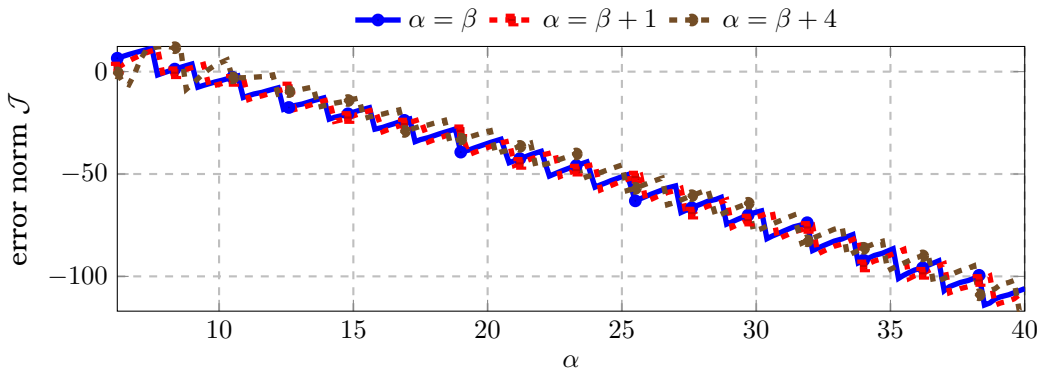


Figure 5.4: Error norm (3.12) for an Jac. diff. with $N = 0$ and cutoff frequency $\omega_c = 500\,\mathrm{rad/s}$ for different values of the parameters α and β.

(2007). When an overdetermined system is used, numerical issues stemming from an ill-conditioning of the matrix Φ are overcome. In the sequel , this approach is called the algebraic approach.

2. Using 2 different Jac. diff. a linear system of equations

$$\boldsymbol{Y} = \Phi\boldsymbol{\theta} \tag{5.21}$$

with $\boldsymbol{Y} = [\hat{y}_1^{(4)}, \hat{y}_2^{(4)}]^T$ and $\Phi = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2]^T$. The gDREM approach[7] is used for the estimation of the parameters. The adaptation gain is $\Gamma = \mathrm{diag}\left(1500\,\mathrm{V}^2/\mathrm{s}, 1500\,\mathrm{V}^2/\mathrm{s}\right)$ and $p = 2$.

The parametrisation of the diffs. is particularly challenging in this context due to the low sampling frequency compared to the highest frequency in the signal. Thus, the bandwidth where the filters g_i act as diffs. must be large enough to ensure that the component from ω_2 is not damped too much, and the discretisation should not alter the continuous-time properties. Jacobi diffs. with $N = 0$, cutoff frequencies equal to $\omega_c = 500\,\mathrm{rad/s}$, and different values of α and β are considered in the sequel. Fig. 5.4 shows the variation of the error norm defined in (3.12) for a Jac. diff. Thus, to get an error norm less than $-95\,\mathrm{dB}$ the parameters α and β must be larger than 35.

Systems of equations of the form (5.20) and (5.21) are now generated using sufficiently many Jac. diffs. with cutoff frequency[8] $\omega_c = 500\,\mathrm{rad/s}$ and different values of α and β. Table 5.1 summarises the properties[9] of the diffs. used. For the gDREM approach or the algebraic approach with two equations the diffs. $g_{0,\vartheta,T}^{(35,36)}$ and $g_{0,\vartheta,T}^{(36,35)}$ are used. This results in two filters with a window length of $22t_s$ and corresponds to 21 filter coefficients when the mid-point rule

[7] In Aranovskiy et al. (2016a) the DREM approach is used for the estimation of the frequencies of a linear combination of two harmonic signals. State variable filters (see, e.g., Peter and Isermann (1990); Young (1981)) are used for the computation of the derivatives of the signal y. New linearly independent equations are generated using pure delay operators. An upper bound on the frequencies in the signal y needs to be known for the choice of the delays. The use of the state variable filters induces an exponentially decaying error term in the estimation of the derivatives as opposed to the approach followed here.

[8] Choosing the cutoff frequency of the Jac. diff. and varying the parameters α and β rather than fixing the window length T and varying α and β seems more intuitive to the author. Engineers generally have some basic knowledge of the system and the frequencies to be expected in the signals. Thus, a cutoff frequency can be chosen based on this knowledge.

[9] Even though the diffs. have been parametrised such that the cutoff frequency is the same for all of them, the chosen parametrisations result in filters with equal window lengths. This equality is due to the low sampling rate. The reader should be aware that this is no longer the case for a higher sampling rate. For example, choosing the sampling period as $t_s = 0.1\,\mathrm{ms}$ yields for $g_{0,\vartheta,T}^{(40,35)}$ and $g_{0,\vartheta,T}^{(38,37)}$ windows of lengths $1151t_s$ and $1128t_s$, respectively.

Table 5.1: Jac. diffs. used for the generation of new equations in the first experimental case study and the value of the error norm (3.12) quantifying the effects of the discretisation for $\omega_\mathrm{c} = 500\,\mathrm{rad/s}$.

Approach	diffs.	error norm (3.12) in dB	T
gDREM and Alg. $N_\mathrm{eq} = 2$	$g_{0,\vartheta,T}^{(35,36)}$	-99.324	$22t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(36,35)}$	-99.324	$22t_\mathrm{s}$
Alg. $N_\mathrm{eq} = 3$	$g_{0,\vartheta,T}^{(35,37)}$	-96.274	$22t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(36,36)}$	-96.600	$22t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(37,35)}$	-96.274	$22t_\mathrm{s}$
Alg. $N_\mathrm{eq} = 6$	$g_{0,\vartheta,T}^{(35,40)}$	-101.819	$23t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(36,39)}$	-103.075	$23t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(37,38)}$	-103.721	$23t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(38,37)}$	-103.721	$23t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(39,36)}$	-103.075	$23t_\mathrm{s}$
	$g_{0,\vartheta,T}^{(40,35)}$	-101.819	$23t_\mathrm{s}$

is used for the discretisation. The parameters α and β are chosen such that the error norm (3.12) is lower than $-95\,\mathrm{dB}$. For the algebraic approach, the scenarios with $N_\mathrm{eq} = 3$ and $N_\mathrm{eq} = 6$ are also considered.

Exploiting the FIR interpretation of the Jac. diffs. and the frequency-domain analysis it can be shown after some tedious but straightforward trigonometric manipulations that $t \mapsto \det\{\Phi(t)\} \notin \mathcal{L}_2$. Thus, the gDREM approach will converge. Fig. 5.5 shows the time evolution of the estimates of the frequencies using the gDREM approach and the algebraic one with $N_\mathrm{eq} = 6$. Since the window length of the Jac. diff. for the algebraic approach is $23t_\mathrm{s}$ it is clear that the approach will converge in finite time to the true value (except for the effects of the measurement disturbance) at $t = 0.115\,\mathrm{s}$. Contrarily, the gDREM estimates converge only asymptotically. However, from the figure it is clear that the convergence is very fast. Both estimates show qualitatively good results, while the ones from the gDREM show slightly less fluctuation around the true value.

The normalised root-mean-square error

$$\text{NRMS error } \omega_i = \sqrt{\sum_{k=k_0}^{N_\mathrm{s}} \frac{\left(\omega_i - \hat{\omega}_i(kt_\mathrm{s})\right)^2}{\omega_i}}, \quad i \in \{1, \ldots, 2\}, \tag{5.22}$$

with k_0 the index corresponding to $t = 0.15\,\mathrm{s}$ and N_s the total number of measurements, is used for the quantitative comparison of the results. Table 5.2 summarises its values for the different approaches. The table confirms the observations from Fig. 5.5. The estimates from the gDREM have a root-mean-square error that is lower than those from the algebraic approach. Moreover, increasing N_eq decreases the error.

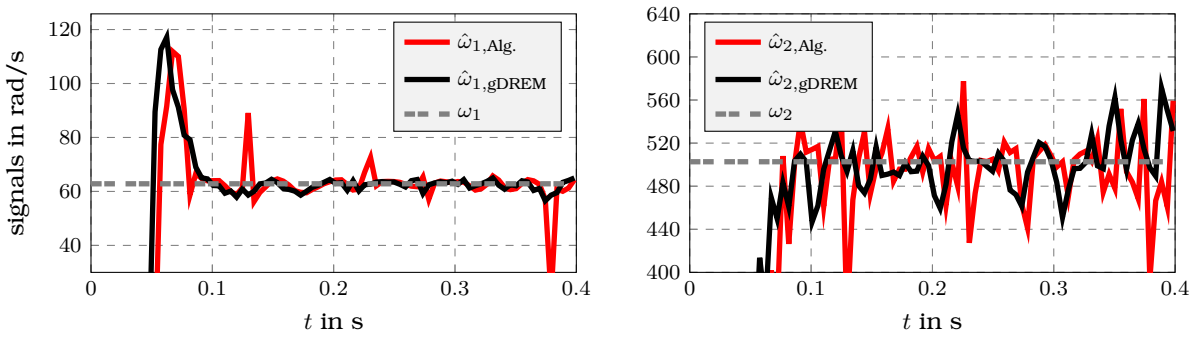


Figure 5.5: Time evolution of the estimates for the experimental case study 1. For the algebraic approach $N_{\text{eq}} = 6$.

Table 5.2: Normalised root-mean-square errors from (5.22) for the first experimental case study.

	gDREM	Alg. $N_{\text{eq}} = 2$	Alg. $N_{\text{eq}} = 3$	Alg. $N_{\text{eq}} = 6$
NRMS error for ω_1	0.0170	0.214	0.129	0.097
NRMS error for ω_2	0.0282	0.444	0.154	0.062

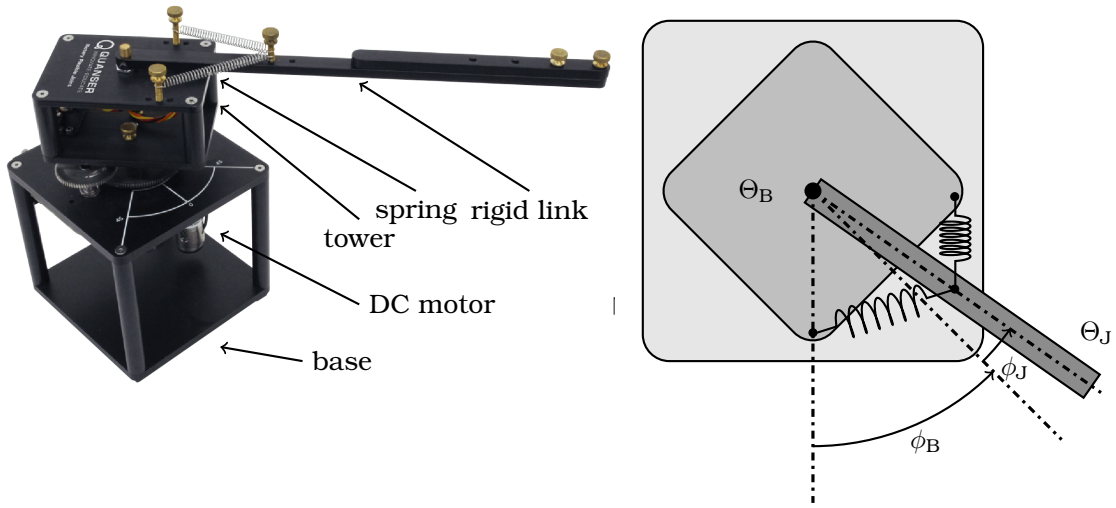


Figure 5.6: Experimental setup of the flexible joint system[11] and a schematic representation.

5.3.3 Experimental case study 2

A second experimental case study for the validation of the parameter estimation approach is performed using a rotary flexible joint system[10] from Quanser. Fig. 5.6 shows the experimental setup and a schematic representation of it. The system consist of a rigid link mounted on a flexible joint. An encoder measures the deflection of the joint whose rotation in the horizontal plane is counteracted by two extension springs. The joint is attached to a tower the rotation of which is driven by a DC motor though some gears. The motor and an encoder measuring the angle of the tower are in the fixed base.

[10] See https://www.quanser.com/products/rotary-flexible-joint/ (accessed: 31:07:2021) for more details.
[11] Photo of flexible joint system from https://www.quanser.com/products/rotary-flexible-joint/ (accessed: 31.07.2021)

Mathematical model of the system

Denote in the following by ϕ_{B} the angle of the tower with respect to some reference as shown in Fig. 5.6. The deflection of the joint with respect to a fixed reference on the tower is denoted by ϕ_{J}. A viscous damping with parameter d is assumed for the joint and the springs shall be approximated by a linear torsion spring with stiffness c_{T}. Neglecting the electrical dynamics of the system and applying the Lagrange formalism yield the differential equations

$$\ddot{\phi}_{\mathrm{J}} = -\Theta c\phi_{\mathrm{J}} - \Theta D\dot{\phi}_{\mathrm{J}} + K\dot{\phi}_{\mathrm{B}} - bu \tag{5.23a}$$

$$\ddot{\phi}_{\mathrm{B}} = c\phi_{\mathrm{J}} + D\dot{\phi}_{\mathrm{J}} - K\dot{\phi}_{\mathrm{B}} + bu \tag{5.23b}$$

with u the voltage applied to the DC motor. In the latter equations the parameters are:

$$\Theta = 1 + \frac{\Theta_{\mathrm{B}}}{\Theta_{\mathrm{J}}}, \quad D = \frac{d}{\Theta_{\mathrm{B}}}, \quad c = \frac{c_{\mathrm{T}}}{\Theta_{\mathrm{B}}}, \quad K = \frac{K_u^2}{\Theta_{\mathrm{B}} R}, \quad b = \frac{K_u}{\Theta_{\mathrm{B}} R},$$

with Θ_{B} and Θ_{J} inertia parameters of the tower and the link, respectively, K_u the motor constant, and R the electrical resistance in the motor. The knowledge of the parameters Θ_{B}, Θ_{J}, R, K_u, d, c_{T} is not required for the design of controllers or observers. Only the numerical values of Θ, c, D, K, and b are required. Their estimation is now discussed.

Parameter estimation

Let $g = g_{0,T,\vartheta}^{(\alpha,\beta)}$ be a Jac. diff. with $\alpha, \beta > 1$ defined in Definition 2.8 and denote by

$$\hat{f}^{(n)}(t) = \int_0^T g^{(n)}(\tau) f(t-\tau) \mathrm{d}\tau$$

the approximation of the n-th order derivative of some square integrable function f. Convolving both sides of (5.23a) and (5.23b) with g yields

$$\ddot{\hat{\phi}}_{\mathrm{J}} = -\Theta c\hat{\phi}_{\mathrm{J}} - \Theta D\dot{\hat{\phi}}_{\mathrm{J}} + K\dot{\hat{\phi}}_{\mathrm{B}} - b\hat{u} \tag{5.24a}$$

$$\ddot{\hat{\phi}}_{\mathrm{B}} = c\hat{\phi}_{\mathrm{J}} + D\dot{\hat{\phi}}_{\mathrm{J}} - K\dot{\hat{\phi}}_{\mathrm{B}} + b\hat{u} \tag{5.24b}$$

The simplest approach for the estimation of the parameters (see Remark 5.1 for a different method) is to consider both equations (5.24a) and (5.24b) as two separate LRE

$$\ddot{\hat{\phi}}_{\mathrm{J}} = \boldsymbol{\phi}^T \boldsymbol{\theta}_1 \quad \text{and} \quad \ddot{\hat{\phi}}_{\mathrm{B}} = \boldsymbol{\phi}^T \boldsymbol{\theta}_2 \tag{5.25}$$

with $\boldsymbol{\phi} = [\hat{\phi}_{\mathrm{J}}, \dot{\hat{\phi}}_{\mathrm{J}}, \dot{\hat{\phi}}_{\mathrm{B}}, \hat{u}]^T$, $\boldsymbol{\theta}_1 = [-\Theta c, -\Theta D, K, -b]^T$, and $\boldsymbol{\theta}_2 = [c, D, -K, b]$. The gDREM approach can then be used for the estimation of the components $\theta_{i,j}$ of $\boldsymbol{\theta}_i$, $i \in \{1, 2\}$ and $j \in \{1, \ldots, 4\}$.

Remark 5.1: Second approach for the estimation of the parameters in (5.24)

Using (5.24b) the parameters c, D, K, and b can be estimated using the gDREM approach, for example. Denote by $\hat{c}$ and $\hat{D}$ the resulting estimates of c and D, respectively. The only remaining parameter to be estimated is Θ. Adding (5.24a) and (5.24b) yields

$$\ddot{\hat{\phi}}_{\mathrm{J}} + \ddot{\hat{\phi}}_{\mathrm{B}} - c\hat{\phi}_{\mathrm{J}} - D\dot{\hat{\phi}}_{\mathrm{J}} = \Theta(-c\hat{\phi}_{\mathrm{J}} - D\dot{\hat{\phi}}_{\mathrm{J}}).$$

Replacing now in the latter equation D and c by their corresponding estimates yields

$$\ddot{\hat{\phi}}_{\mathrm{J}} + \ddot{\hat{\phi}}_{\mathrm{B}} - \hat{c}\hat{\phi}_{\mathrm{J}} - \hat{D}\dot{\hat{\phi}}_{\mathrm{J}} = \Theta(-\hat{c}\hat{\phi}_{\mathrm{J}} - \hat{D}\dot{\hat{\phi}}_{\mathrm{J}}),$$

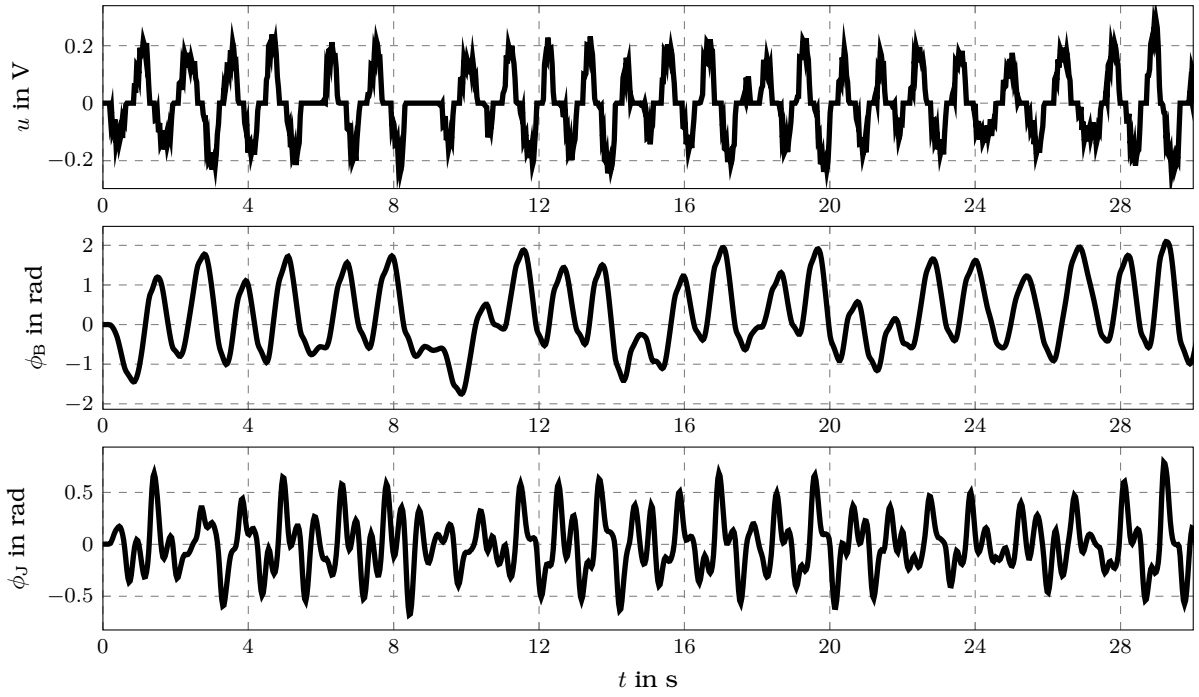


Figure 5.7: Time evolution of the voltage u and the angles ϕ_B and ϕ_J of the identification trajectory.

which is suitable to estimate Θ using again gDREM, for example. Thus, only five parameters are estimated in total as compared to the previous approach where eight are considered.

The measurements of u, ϕ_B, and ϕ_J are sampled at a frequency of $100\,\mathrm{Hz}$. Fig. 5.7 shows the time evolution of these signals that shall be used for the identification of the parameters. A single Jac. diff. with parameters $N = 0$, $\alpha = \beta = 40$, and cutoff frequency $\omega_\mathrm{c} = 250\,\mathrm{rad/s}$ is used in the following. This cutoff frequency corresponds to $80\,\%$ of the Nyquist frequency. Using the mid point rule yields a discrete filter with 25 coefficients. To reduce the computation requirements, the six new equations that are required for the application of gDREM to each of the LRE in (5.25) are created using a pure delay operator with delay $d = 20t_\mathrm{s}$. The parameter p is chosen equal to 1 and the adaptation gains are

$$\Gamma_1 = \Gamma_2 = \mathrm{diag}\,(1, 0.2, 1, 1)\,/\mathrm{sV}$$

for the estimation of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. Note that for this example it is analytically not possible to verify the convergence condition in Proposition 5.2. The results will be verified by comparing measured trajectories with simulated ones. As a comparison the estimates from standard gradient estimators of the form (5.2) with gains

$$\bar{\Gamma}_1 = \bar{\Gamma}_2 = \mathrm{diag}\left(3.2/\mathrm{s}, 0.04\,\mathrm{s}, 0.32\,\mathrm{s}, 36/(\mathrm{s\,V}^2)\right) \tag{5.26}$$

shall be considered. Furthermore, the parameters in $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are also estimated offline as

the least squares solutions of the systems[12]

$$\begin{bmatrix} \ddot{\hat{\phi}}_{\mathrm{J}}(t_1) \\ \vdots \\ \ddot{\hat{\phi}}_{\mathrm{J}}(t_{N_\mathrm{s}}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}^T(t_1) \\ \vdots \\ \boldsymbol{\phi}^T(t_{N_\mathrm{s}}) \end{bmatrix} \boldsymbol{\theta}_1 \quad \text{and} \quad \begin{bmatrix} \ddot{\hat{\phi}}_{\mathrm{B}}(t_1) \\ \vdots \\ \ddot{\hat{\phi}}_{\mathrm{J}}(t_{N_\mathrm{s}}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}^T(t_1) \\ \vdots \\ \boldsymbol{\phi}^T(t_{N_\mathrm{s}}) \end{bmatrix} \boldsymbol{\theta}_2,$$

respectively, with N_s the total number of measurements. Denote in the following by $\theta_{i,j,\mathrm{LS}}$, $i \in \{1,2\}$ and $j \in \{1,\dots,4\}$, the solution of the latter equations.

Fig. 5.8 shows the time evolution of the estimates using the gDREM and the standard gradient approach with the previously discussed parameters. It can be seen that the estimates from the standard gradient from (5.2) with gains given in (5.26) do not converge to the values computed using the offline estimation approach. Increasing the adaptation gains significantly increases the effect of measurement disturbances. In contrast, the estimates from the gDREM approach quickly converge to the values from the offline estimation. Moreover, the exponential convergence can be nicely seen especially for the estimates of the parameters $\theta_{i,j}$, $i \in \{1,2\}$ and $j \in \{2,\dots,3\}$.

Consider now the system

$$\begin{aligned} \ddot{\phi}_{\mathrm{J}} &= \hat{\theta}^*_{1,1}\phi_{\mathrm{J}} + \hat{\theta}^*_{1,2}\dot{\phi}_{\mathrm{J}} + \hat{\theta}^*_{1,3}\dot{\phi}_{\mathrm{B}} + \hat{\theta}^*_{1,4}u \\ \ddot{\phi}_{\mathrm{B}} &= \hat{\theta}^*_{2,1}\phi_{\mathrm{J}} + \hat{\theta}^*_{2,2}\dot{\phi}_{\mathrm{J}} + \hat{\theta}^*_{2,3}\dot{\phi}_{\mathrm{B}} + \hat{\theta}^*_{2,4}u, \end{aligned}$$

with $\hat{\theta}^*_{i,j}$, $i \in \{1,2\}$ and $j \in \{1,\dots,4\}$, the mean values of the last 100 values of estimates from the gDREM approach, for the validation of the approach. This system is simulated with an input trajectory given in Fig. 5.9 and different from that used for the identification and suitable initial conditions. The time evolution of the angles ϕ_{B} and ϕ_{J} from the simulation and the measurement are shown in Fig. 5.9. Comparing the signals proves that the gDREM approach with this Jac. diff. yields very good results.

5.4 Concluding remarks

The estimation of parameters satisfying a LRE has been discussed in this chapter. The DREM approach developed in Aranovskiy et al. (2017) has been recalled, and two modifications have been proposed. These modifications yield less restrictive convergence conditions on the regressor. The second modification ensures that if only a subset of parameters is excited, the adaptation is not entirely stopped as in the standard DREM approach. The advantages of this approach have been discussed using examples. For the generation of new linearly independent equations, the use of Jac. diffs. has been discussed in two experimental case studies. The tuning of the diffs. has been discussed in detail based on the analysis in Chapters 2 and 3. Especially the first example, where derivatives up to the fourth order are required in the presence of a low sampling rate compared to the frequencies in the signals, is challenging. An interesting question for future research is the choice of the parameter p in (5.12) when unmodelled dynamics and disturbances are explicitly considered in the analysis. While choosing $p = 1$ yields less restrictive convergence conditions, as discussed in Example 5.9, no robustness analysis has been carried out.

[12] In Othmane et al. (2020) the errors arising in the estimation of parameters using this approach with Jac. diff. are discussed. In Othmane et al. (2022) an approach for the choice of the parameters of Jac. diff. for this application is discussed in detail.

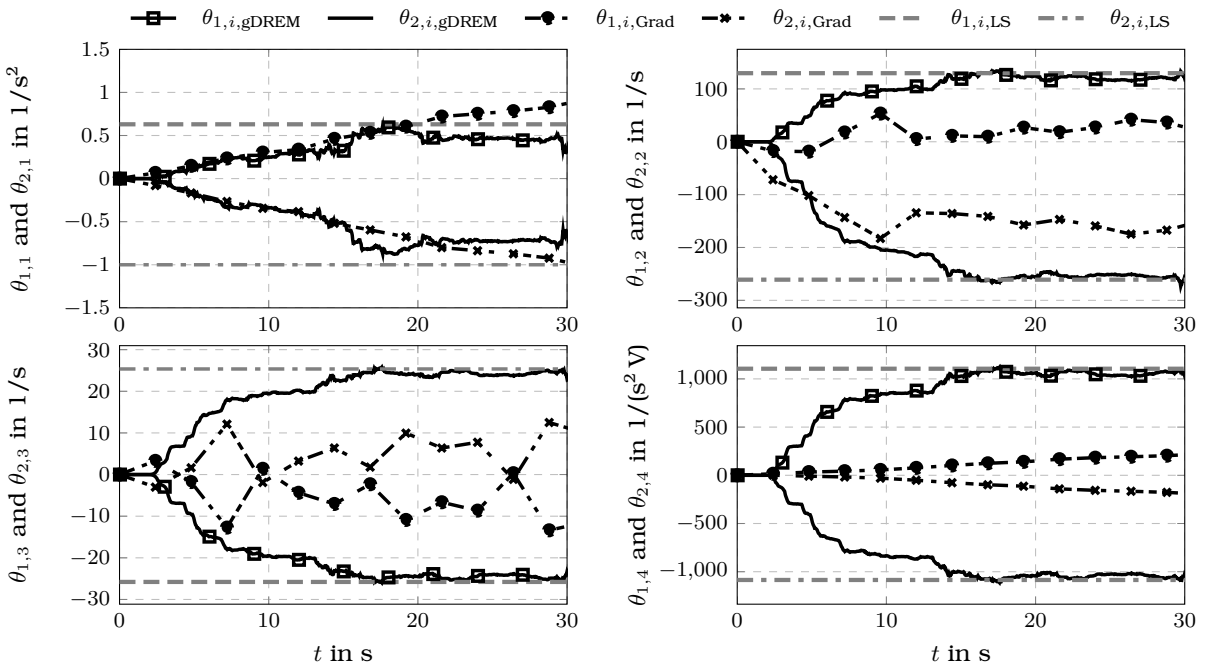


Figure 5.8: Time evolution of the estimates for the flexible joint system for the gDREM approach, the standard gradient approach. The constant lines show the parameter values estimated offline.

During the discussion of the experimental studies, the concept of modulating function has been recalled, and it has been emphasised that Jac. diffs. can be interpreted as modulating functions. This concept is very useful for the estimation of parameters as shown in Preisig and Rippin (1993a, 1993b); Shinbrot (1954, 1957). Modulating functions have been generalised in Jouffroy and Reger (2015); Liu et al. (2014) such that the conditions given in Definition 5.10 are relaxed. The generalisation is recalled in the following.

Definition 5.11: Right and left modulating function (Jouffroy and Reger (2015))

Consider a sufficiently smooth function $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and write one of its partial derivatives as

$$\phi^{(i)}(t, t_1) = \frac{\partial^i \phi}{\partial \tau^i}(\tau, t_1)\Big|_{\tau=t}.$$

The function ϕ is called a modulating function (of order k) if there exists $t_0 < t_1$ such that

$$\phi^{(i)}(t_0, t_1)\phi^{(i)}(t_1, t_1) = 0$$

for all $i \in \{0, \dots, k-1\}$. A modulating function for which $\phi^{(i)}(t_0, t_1) = 0$ and $\phi^{(i)} \neq 0$ is called a left modulating function, while a modulating function for which $\phi^{(i)}(t_0, t_1) \neq 0$ and $\phi^{(i)} = 0$ is called a right modulating function. A modulating function whose boundaries verify $\phi^{(i)}(t_0, t_1) = 0$ and $\phi^{(i)} = 0$ is called a total modulating function.

These types of functions can be used for the finite-time simultaneous estimation of states and parameters of linear (see, e.g., P. Li, Boem, Pin, and Parisini (2018); Li et al. (2020)) systems and special classes of nonlinear polynomial systems (see, e.g., Jouffroy and Reger

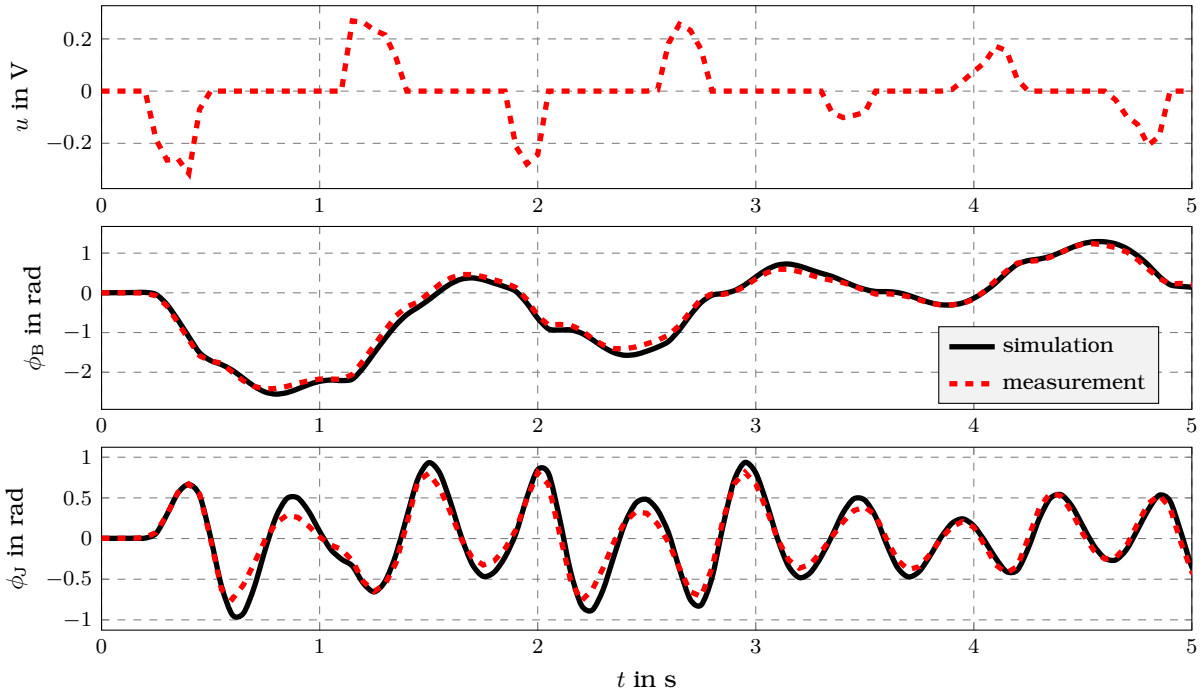


Figure 5.9: Time evolution of the voltage u and the angles ϕ_B and ϕ_J of the measurement and the simulated system.

(2015)). It is now straightforward for the reader to verify the conditions on the parameters α and β for the Jac. diffs. from Definition 2.8 and the parameter α of the Laguerre differentiators (Lag. diffs.) from Definition 2.7 for the filters to act as right and/or left modulating functions. The tuning guidelines applied in the numerous examples from this and the previous chapters can immediately be applied for their parametrisation in the context of simultaneous state and parameter estimation.

6

Concluding remarks and outlook

Approximating the derivatives of measured signals is essential for reconstructing unmeasured quantities. This challenging yet fundamental approximation problem is addressed in this work using methods based on orthogonal polynomials. Chapter 2 discusses differentiators based on the latter functions and provides different interpretations. The analysis of the differentiators shows that the approximated derivatives are solutions to least squares problems. Their algebraic formulation creates a link to parameter estimation problems and their interpretation as linear time-invariant filters is convenient for tuning. While the Jacobi differentiators, also known as algebraic differentiators, have been derived in Mboup et al. (2007); Mboup et al. (2009) and thoroughly discussed in Kiltz (2017); Kiltz and Rudolph (2013); Liu (2011); Mboup and Riachy (2014, 2018); Othmane et al. (2022), those based on Hermite and Laguerre polynomials are discussed in more detail here than in the literature known to the author. In particular, the approximation error, delay, concrete tuning guidelines, and implementation issues are investigated in this work. In the opinion of the author, the here used approximation-theoretical derivation of the differentiators allows for a more illustrative interpretation of the methods and their parameters than the abstract differential-algebraic approach used in Mboup et al. (2007); Mboup et al. (2009); Ushirobira (2018); Ushirobira and Quadrat (2016), for instance.

Systematic tuning guidelines are investigated in Chapter 3 for differentiators based on classical orthogonal polynomials. Those based on Hermite, Laguerre, and Jacobi polynomials depend on three, four, and five parameters, respectively, and their choice is essential to achieve satisfying approximations. It is shown that the amplitude spectra of Laguerre differentiators can be well approximated by a simple piecewise function, if a known delay is accepted. This has already been shown for Jacobi differentiators in Kiltz (2017); Kiltz and Rudolph (2013). These simple yet powerful approximations yield straightforward tuning guidelines: The parameters of the differentiators can be computed from desired cutoff frequencies and stopband slopes. Section 3.2.2 shows that high-gain differentiators and state variable filters can be recovered from Laguerre differentiators for particular parameter choices. The guidelines discussed are crucial to find suitable parametrisations without relying on trial-and-error approaches and allow their use in numerous control engineering applications. The application of the methods for Jacobi differentiators in the context of model-free control has been considered in Othmane, Rudolph, and Mounier (2021a), for examples. The use of Laguerre differentiators in this context should be considered in future research. The publicly available toolbox from Othmane (2021) for the design, analysis, and discretization of Jacobi

differentiators and incorporating the here discussed tuning guidelines significantly simplifies the use of these filters.

In Section 3.3, the developed tuning guidelines are applied in concrete experimental case studies. These discussions show that convincing results can be achieved using all the differentiators, especially when slight delays are tolerable. Comparisons with high-gain and sliding mode differentiators have already been published in Liu (2011); Othmane, Rudolph, and Mounier (2021b); Sidhom (2011), for example, where it has been shown that Jacobi differentiators outperform the other two approaches. In the case study from Section 3.3, the Hermite differentiator requires a larger filter window length than the ones based on Jacobi and Laguerre polynomials to achieve similar noise rejection properties. Thus, a real-time implementation of the former yields a higher computation burden. The experimental results presented in Section 3.3.2 shows that the differentiators can be efficiently used for the approximate inversion of analogue low-pass filters.

Differentiators based on Jacobi polynomials are FIR filters. In contrast, the implementation of those based on Hermite and Laguerre polynomials requires their approximation as FIR filters. The distribution function (3.9) is used for the computation of the filter window length. In contrast, the estimates of Laguerre differentiators having integer orders can be recovered by the outputs of stable LTI systems, which considerably reduces the computational burden. This is a significant advantage for application scenarios using a low-cost hardware. A further difference between the differentiators can be observed in the phase of their Fourier transform. Jacobi differentiators can have a linear phase for specific parametrisations as shown in (Kiltz, 2017, Sec. 3.3.1). This property can be a significant advantage. In contrast, neither Hermite nor Laguerre differentiators can be parametrised as linear phase filters. Thus, multiple trade-offs must be considered in designing the differentiators. Despite the numerous tuning guidelines, the final parametrisation of these filters will still most likely be made by engineering judgments on the specifications regarding the frequency-domain properties, method of implementation, computational burden, and memory facilities.

A further simplification of the tuning process is proposed in Section 3.4. The optimisation problem (3.30) is based only on the measured signal and takes into account discretisation effects. The parametrisation problem is boiled down to the choice of a single parameter. The computed filter can then be used for the approximation of derivatives. While Section 3.4.2 shows promising results for an experimental case study, numerous theoretical questions remain open and require further investigations. The choice of the optimisation algorithms needs to be considered in detail. An analysis of the effects of the different parameters on the approximated derivative has to be addressed. A further interesting extension of the approach is the design of adaptive filters that automatically select optimal parameters to maximise the approximation accuracy. The computation of new parameters can then be automatically triggered, if the quantity (3.29) exceeds a given threshold. An adaptive differentiation filter for tracking instantaneous frequencies in power systems has already been developed in Zivanovic (2007), for example. The proposed algorithm automatically adapts the filter window length to maximize measurement accuracy. Large values are chosen if the considered signal has slow dynamics. In contrast, the length is reduced for fast-varying ones. However, the approach is limited to the estimation of first-order derivatives. Thus, extending the ideas from Section 3.4 seems promising. The tuning approach can be further developed to incorporate a maximally tolerable delay as discussed in Remark 3.1, for example. Moreover,

the group delay (see, e.g., Papoulis (1962)) of differentiators and its variation should be small to reduce the estimation delay and phase distortion. This can be incorporated into the tuning approach by including in the cost function of (3.30) the arithmetic mean of the group delay and its variance over a set of relevant frequencies. The application considered in Section 4.1 requires the differentiators to attenuate a known but undesired frequency stemming from a mechanical eigenmode. By adapting the cost function such that it includes the amplitude of the Fourier transform of the differentiator evaluated in an interval of interest it can be ensured that the undesired frequency is sufficiently attenuated. Alternatively, the ratio of this amplitude and that evaluated at the cutoff frequency can be considered.

Differentiators based on Jacobi and Laguerre polynomials are used in Chapter 4 for the model-based detection of impulsive forces. In Section 4.1, it is shown that the collision of a table tennis ball with a plate can be efficiently detected. Combining a Laguerre differentiator with a notch filter requires less computational burden and yields a lower minimum detectable falling height than the approach based on Jacobi differentiators from Kiltz (2017); Kiltz and Rudolph (2013). Other application scenarios for the developed approach might be fault detection and identification on the same experimental setup to implement fault-tolerant control algorithms as discussed in Kiltz et al. (2014, 2012), for example. In Section 4.2, Jacobi differentiators are used for the model-based detection of faults in rolling element bearings. The potential of the approach is validated through experiments with several publicly available data sets. While in the considered experimental case studies the shaft is rotating with a constant velocity, the approach is not limited to this case. Investigating the effectiveness of the method for systems with time-varying velocities should be considered in future works.

The ideas developed for parameter estimation problems in Chapter 5 extend previously published works (see, e.g., Aranovskiy et al. (2017); Ortega et al. (2021); Othmane et al. (2020)). Their experimental validation and the comparison with other approaches show promising results. However, numerous open questions have to be considered. Even if the analysed examples contain disturbed measurements and the approach yields convincing results, the effects of disturbances has yet to be analysed systematically. Moreover, the tuning parameter p used in Section 5.2 may be considered in more detail. Especially the robustness of the approach with respect to unmodelled dynamics and the influence of p require further investigations.

A

Special functions and their properties

This chapter lists the properties of special functions used in this work. For a detailed discussion the reader is referred to Abramowitz and Stegun (1965); Lebedev and Silverman (1965); Szegö (1939); Watson (1995).

Chapter content

A.1 Classical orthogonal polynomials 141
A.1.1 Hermite Polynomials 141
A.1.2 Laguerre polynomials 142
A.1.3 Jacobi polynomials 143
A.2 Other functions 143
A.2.1 Gamma function and related functions 143
A.2.2 Error function 143
A.2.3 Beta function and related functions: 144
A.2.4 Sinc function 144
A.2.5 Bessel function of the first kind 144

A.1 Classical orthogonal polynomials

A.1.1 Hermite Polynomials

#A.1 The explicit expression of Hermite polynomials is

$$\mathrm{H}_n(\tau) = n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m 2^{n-2m}}{m!(n-2m)!} \tau^{n-2m},$$

with $\lfloor x \rfloor = \max_{z \in \mathbb{Z}} z \leqslant x$ the floor function and $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ the set of integers.

#A.2 Hermite polynomials are even or odd functions and thus satisfy the symmetry property

$$\mathrm{H}_n(-\tau) = (-1)^n \mathrm{H}_n(\tau).$$

#A.3 Recalling that

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\mathrm{e}^{-\tau^2}\mathrm{H}_m(\tau)\right) = -\mathrm{e}^{-\tau^2}\mathrm{H}_{m+1}(\tau)$$

yields

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^n\left(\mathrm{e}^{-\tau^2}\mathrm{H}_m(\tau)\right) = (-1)^n\mathrm{e}^{-\tau^2}\mathrm{H}_{m+n}(\tau).$$

#A.4 From #A.3 it follows that

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^n\mathrm{e}^{-\tau^2} = (-1)^n\mathrm{e}^{-\tau^2}\mathrm{H}_n(\tau).$$

#A.5 The Hermite polynomials satisfy

$$\int_0^\tau w(\sigma)\mathrm{H}_n(\sigma)\mathrm{d}\sigma = \mathrm{H}_{n-1}(0) - w(\tau)\mathrm{H}_n(\tau) \tag{A.1a}$$

for $n > 0$ and

$$\int_0^\tau w(\sigma)\mathrm{H}_0(\sigma)\mathrm{d}\sigma = \frac{\sqrt{\pi}}{2}\mathrm{erf}\,(\tau)\,, \tag{A.1b}$$

with $\tau \mapsto \mathrm{erf}\,(\tau)$ the error function defined in #A.15.

A.1.2 Laguerre polynomials

#A.6 The explicit expression of a Laguerre polynomial denoted by $\mathrm{L}_n^{(\alpha)}$, with $\alpha > -1$ and $n \in \mathbb{N}$, is

$$\mathrm{L}_n^{(\alpha)}(\tau) = \sum_{m=0}^{n}\frac{(-1)^m}{m!}\binom{n+\alpha}{n-m}\tau^m, \quad \binom{n+\alpha}{n-m} = \frac{\Gamma(n+\alpha+1)}{\Gamma(n-m+1)\Gamma(\alpha+m+1)}.$$

#A.7 The product[1] $\mathrm{L}_n^{(\alpha)} \cdot w_\alpha$, with $w_\alpha(\tau) = \mathrm{e}^{-\tau}\tau^\alpha$, can be written as

$$w_\alpha(\tau)\mathrm{L}_n^{(\alpha)}(\tau) = \sum_{m=0}^{n}\frac{(-1)^m}{m!}\binom{n+\alpha}{n-m}w_{\alpha+m}.$$

#A.8 Recalling that

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\mathrm{e}^{-\tau}\tau^\alpha\mathrm{L}_m^{(\alpha)}(\tau)\right) = (m+1)\mathrm{e}^{-\tau}\tau^{\alpha-1}\mathrm{L}_{m+1}^{(\alpha-1)}(\tau)$$

yields

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^n\left(\mathrm{e}^{-\tau}\tau^\alpha\mathrm{L}_m^{(\alpha)}(\tau)\right) = \frac{(m+n)!}{m!}\mathrm{e}^{-\tau}\tau^{\alpha-n}\mathrm{L}_{m+n}^{(\alpha-n)}(\tau).$$

for $\alpha > n-1$.

#A.9 The property in #A.8 yields

$$\begin{aligned}\int_0^t \mathrm{e}^{-\tau}\tau^{\alpha-n}\mathrm{L}_{m+n}^{(\alpha-n)}(\tau)\mathrm{d}\tau &= \frac{m!}{(m+n)!}\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\right)^{n-1}\left(\mathrm{e}^{-\tau}\tau^\alpha\mathrm{L}_m^{(\alpha)}(\tau)\right)\Big|_{\tau=t}\\ &= \frac{1}{(m+n)}\mathrm{e}^{-t}t^{\alpha-n+1}\mathrm{L}_{m+n-1}^{(\alpha-n+1)}(t)\end{aligned}$$

[1] The pointwise product of two functions f and g is denoted by $f \cdot g$.

#A.10 Let $g^{(\alpha)} : \tau \mapsto g^{(\alpha)}(\tau) = \mathrm{e}^{-\tau}\tau^{\alpha}$ and $\alpha \in \mathbb{N}_{>1}$. A proof by induction yields

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau} + 1\right)^{\alpha+1} g^{(\alpha)}(\tau) = 0.$$

It then follows that the function $f : \tau \mapsto f(\tau) = \sum_{n=0}^{N} c_n g^{(\alpha+n)}(\tau)$ satisfies the differential equation

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau} + 1\right)^{\alpha+N} f(\tau) = c_N(N+\alpha)!\mathrm{e}^{-\tau}.$$

or equivalently

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tau} + 1\right)^{\alpha+N+1} f(\tau) = 0.$$

A.1.3 Jacobi polynomials

#A.11 The explicit representation of a Jacobi polynomial $\mathrm{P}_n^{(\alpha,\beta)}$, with $\alpha, \beta > -1$, is

$$\mathrm{P}_n^{(\alpha,\beta)}(\tau) = \sum_{m=0}^{n} = \frac{(-1)^{n-m}}{2^n}\binom{n+\alpha}{m}\binom{n+\beta}{n-m} w^{(n-m,m)}(\tau)$$

with

$$w^{(\alpha,\beta)}(\tau) = (1-\tau)^{\alpha}(1+\tau)^{\beta}.$$

#A.12 The n-th order derivative of the product $w^{(\alpha,\beta)} \cdot \mathrm{P}_n^{(\alpha,\beta)}$ is

$$\left(w^{(\alpha,\beta)} \cdot \mathrm{P}_m^{(\alpha,\beta)}\right)^{(n)} = \frac{(-2)^n(n+m)!}{m!} w^{(\alpha-n,\beta-n)} \cdot \mathrm{P}_{n+m}^{(\alpha-n,\beta-n)}.$$

A.2 Other functions

A.2.1 Gamma function and related functions

#A.13 The Gamma function $\Gamma : z \mapsto \Gamma(z)$ is defined as

$$\Gamma(z) = \int_0^{\infty} \tau^{z-1}\mathrm{e}^{-\tau}\mathrm{d}\tau, \quad z \in \mathbb{C},$$

with $\mathcal{R}\{z\} > 0$, where $\mathcal{R}\{z\}$ denotes the real part of z. For a natural number $\alpha \in \mathbb{N}$ it simplifies to $\Gamma(\alpha) = (\alpha-1)!$. The Gamma function can be used to define a binomial coefficient as

$$\binom{\alpha}{\beta} = \frac{\Gamma(\alpha+1)}{\Gamma(\beta+1)\Gamma(\alpha-\beta+1)}.$$

#A.14 The lower incomplete gamma function $\gamma : (\alpha, t) \mapsto \gamma(\alpha, t)$ is defined as

$$\gamma(\alpha, t) = \int_0^{t} \tau^{\alpha-1}\mathrm{e}^{-\tau}\mathrm{d}\tau.$$

for $t \in \mathbb{C}$ and $a \in \mathbb{C}$ with $\mathcal{R}\{\alpha\} > 0$. From #A.13 it follows that $\lim_{t\to\infty} \gamma(\alpha, t) = \Gamma(\alpha)$.

A.2.2 Error function

#A.15 The error function $\mathrm{erf} : \tau \mapsto \mathrm{erf}(\tau)$ is defined as

$$\mathrm{erf}(\tau) = \frac{2}{\sqrt{\pi}} \int_0^{\tau} \mathrm{e}^{-\sigma^2}\mathrm{d}\sigma.$$

A.2.3 Beta function and related functions:

#A.16 The beta function $\mathrm{B} : (\alpha, \beta) \mapsto \mathrm{B}(\alpha, \beta)$ is defined as

$$\mathrm{B}(\alpha, \beta) = \int_0^1 \tau^{\alpha-1}(1-\tau)^{\beta-1}\mathrm{d}\tau = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

For natural numbers α and β it simplifies to

$$\mathrm{B}(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)}.$$

#A.17 The incomplete beta function $\mathrm{B}_\tau : (\alpha, \beta) \mapsto \mathrm{B}_\tau(\alpha, \beta)$ is defined as

$$\mathrm{B}_\tau(\alpha, \beta) = \int_0^\tau \sigma^{\alpha-1}(1-\sigma)^{\beta-1}\mathrm{d}\sigma.$$

#A.18 The regularised incomplete beta function $I_\tau : (\alpha, \beta) \mapsto I_\tau(\alpha, \beta)$ is defined as

$$I_\tau(\alpha, \beta) = \frac{\mathrm{B}_\tau(\alpha, \beta)}{\mathrm{B}(\alpha, \beta)}.$$

A.2.4 Sinc function

#A.19 The sinc function $x \mapsto \mathrm{sinc}(x)$ is defined as

$$\mathrm{sinc}(x) = \begin{cases} 1, & \text{for} \quad x = 0 \\ \frac{\sin x}{x}, & \text{otherwise.} \end{cases}$$

The function satisfies

$$\int_0^\infty \mathrm{sinc}(\tau)^2\,\mathrm{d}\tau = \int_0^\infty \mathrm{sinc}(\tau)^2\,\mathrm{d}\tau = \frac{\pi}{2},$$

as derived in (Courant & Hilbert, 1924, P. 55) or (Apostol, 1967, P. 286 and 300), for example. However, the integral

$$\int_0^\infty |\mathrm{sinc}(\tau)|\,\mathrm{d}\tau$$

diverges as discussed in (Bartle & Sherbert, 2000, P. 297).

A.2.5 Bessel function of the first kind

#A.20 The Bessel function of the first kind and order $\nu > 0$ is defined as

$$\mathrm{J}_\nu(z) = \sum_{k=0}^\infty \frac{(-1)^k}{\Gamma(\nu+1+k)}\left(\frac{z}{2}\right)^{2k+\nu}$$

for $z \in \mathbb{C}$. The function possesses infinitely many zeros $\mathrm{j}_{\nu,i}$, $i \in \mathbb{N}$, with $0 < \mathrm{j}_{\nu,0} < \mathrm{j}_{\nu,1} < \ldots$

B

The Fourier transform and its properties

This chapter recalls the properties of the Fourier transform (see, e.g., Papoulis (1962) for more details). Furthermore, the Fourier transform of classical OPs and related functions are recalled. The reader is referred to Abramowitz and Stegun (1965); Erdélyi, Magnus, Oberhettinger, and Tricomi (1954b) for more details.

Content

B.1 Properties of the Fourier transform 145
B.2 Fourier transform of special functions 146

B.1 Properties of the Fourier transform

#B.5 The Fourier transform $\mathcal{F}\{f\}$ of a function f is defined as

$$\mathcal{F}\{f\}(\omega) = F(\omega) \coloneqq \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\iota\omega t}\mathrm{d}t, \quad \iota^2 = -1.$$

The inverse of the latter transformation is

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\mathrm{e}^{-\iota\omega t}\mathrm{d}\omega.$$

#B.6 The shifting property is

$$f(t-\mu) \leftrightarrow \mathrm{e}^{-\iota\omega\mu}\mathcal{F}\{f\}(\omega).$$

#B.7 The time scaling property is

$$f(tT) \leftrightarrow \frac{1}{|T|}\mathcal{F}\{f\}\left(\frac{\omega}{T}\right).$$

#B.8 Combining #B.6 and #B.7 yields

$$f\left(\frac{t-\mu}{T}\right) \leftrightarrow \mathrm{e}^{-\iota\omega\mu}|T|\mathcal{F}\{f\}(T\omega).$$

#B.9 The Fourier transform of a multiplication of a function with a monomial is

$$t^n f(t) \leftrightarrow \left(\iota\frac{\mathrm{d}}{\mathrm{d}\omega}\right)^n \mathcal{F}\{f\}(\omega).$$

B.2 Fourier transform of special functions

B.2.1 Hermite polynomials

#B.10 The Fourier transform of the weight function $w : \tau \mapsto w(\tau) = \mathrm{e}^{-\tau^2}$ is

$$\mathcal{F}\{w\}(\omega) = \sqrt{\pi}\mathrm{e}^{-\frac{\omega^2}{4}}.$$

#B.11 The Fourier transform of $w \cdot \mathrm{H}_n$ can be computed using the property #B.9 and the explicit definition of the Hermite polynomial H_n given in #A.1. It follows that

$$\mathcal{F}\{w \cdot \mathrm{H}_n\}(\omega) = n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m 2^{n-2m}}{m!(n-2m)!} \left(\iota \frac{\mathrm{d}}{\mathrm{d}\omega}\right)^{n-2m} \mathcal{F}\{w\}(\omega).$$

Using the higher order derivative of the weight function given in #A.4 the Fourier transform becomes

$$\begin{aligned}\mathcal{F}\{w \cdot \mathrm{H}_n\}(\omega) &= n!\sqrt{\pi} \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{m!(n-2m)!} (-\iota)^{n-2m} \mathrm{e}^{-\frac{\omega^2}{4}} \mathrm{H}_{n-2m}\left(\frac{\omega}{2}\right) \\ &= (-\iota)^n \sqrt{\pi}\mathrm{e}^{-\frac{\omega^2}{4}} \omega^n.\end{aligned}$$

Combining this results with the property #B.8 yields

$$\mathcal{F}\{(w \cdot \mathrm{H}_n) \circ \theta_T\}(\omega) = T\sqrt{\pi}\mathrm{e}^{-\iota\omega\mu} w\left(\frac{T\omega}{2}\right)(\iota T\omega)^n,$$

for $\tau \mapsto \theta_T(\tau) = \frac{\mu-\tau}{T}$, for $T > 0$.

B.2.2 Laguerre polynomials

#B.12 The Fourier transform of the function

$$w_\alpha : \tau \mapsto w_\alpha(\tau) = \begin{cases} \tau^\alpha \mathrm{e}^{-\tau}, & \text{for} \quad \tau \geqslant 0, \\ 0, & \text{otherwise} \end{cases}$$

for $\alpha > -1$, is

$$\mathcal{F}\{w_\alpha\}(\omega) = \int_{-\infty}^{\infty} w(t)\mathrm{e}^{-\iota\omega t}\mathrm{d}t = \int_0^\infty t^\alpha \mathrm{e}^{-(\iota\omega+1)t}\mathrm{d}t = \frac{\Gamma(\alpha+1)}{(1+\iota\omega)^{\alpha+1}},$$

with Γ the Gamma function from #A.13.

#B.13 Using the explicit definition of a Laguerre polynomial $\mathrm{L}_n^{(\alpha)}$ given in #A.6 and the Fourier transform from #B.12 it can be shown that

$$\mathcal{F}\left\{w_\alpha \cdot \mathrm{L}_n^{(\alpha)}\right\} = \frac{\Gamma(\alpha+n+1)(\iota\omega)^n}{n!(1+\iota\omega)^{\alpha+n+1}}.$$

References

Abel, N. H. (1826). Untersuchungen über die Reihe: $1+\frac{m}{1}x+\frac{m\cdot(m-1)}{1\cdot 2}\cdot x^2+\frac{m\cdot(m-1)\cdot(m-2)}{1\cdot 2\cdot 3}\cdot x^3+\ldots$. *J. Reine Angew. Math.*, *1*, 311-339. doi: 10.1515/crll.1826.1.311

Abramowitz, M., & Stegun, I. A. (Eds.). (1965). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover.

Adams, R. A., & Fournier, J. J. F. (1975). *Sobolev spaces*. New York, San Francisco, London: Academic Press.

Ahmed, H., Ushirobira, R., Efimov, D., Tran, D., Sow, M., Payton, L., & Massabuau, J. (2016). A fault detection method for automatic detection of spawning in oysters. *IEEE Trans. Control Syst. Technol.*, *24*(3), 1140-1147. doi: 10.1109/TCST.2015.2472999

Ahnert, K., & Abel, M. (2007). Numerical differentiation of experimental data: local versus global methods. *Comput. Phys. Commun.*, *177*(10), 764-774. doi: 10.1016/j.cpc.2007.03.009

Alavi, S. M., & Saif, M. (2010). A decentralized technique for robust simultaneous fault detection and control of uncertain systems. In *Proc. of the 2010 American Control Conf.* (p. 5445-5450). doi: 10.1109/ACC.2010.5530795

Anderssen, R. S., & Bloomfield, P. (1974). Numerical differentiation procedures for non-exact data. *Numer. Math.*, *22*(3), 157-182. doi: 10.1007/bf01436965

Anderssen, R. S., & de Hoog, F. R. (1984). Finite difference methods for the numerical differentiation of non-exact data. *Computing*, *33*(3-4), 259-267. doi: 10.1007/bf02242272

Anderssen, R. S., & Hegland, M. (1999). For numerical differentiation, dimensionality can be a blessing! *Math. Comput.*, *68*(227), 1121-1141. doi: 10.1090/s0025-5718-99-01033-9

Apostol, T. M. (1967). *Calculus, volume 1*. New York, Santa Barbara, London, Sydney, Toronto: John Wiley & Sons.

Aranovskiy, S., Bobtsov, A., Ortega, R., & Pyrkin, A. (2016b). Parameters estimation via dynamic regressor extension and mixing. In *Proc. of the 2016 American Control Conf.* (p. 6971-6976). Boston, USA. doi: 10.1109/ACC.2016.7526771

Aranovskiy, S., Bobtsov, A., Ortega, R., & Pyrkin, A. (2017). Performance enhancement of parameter estimators via dynamic regressor extension and mixing. *IEEE Trans. Automat. Contr.*, *62*(7), 3546-3550. doi: 10.1109/TAC.2016.2614889

Aranovskiy, S., Bobtsov, B., Ortega, R., & Pyrkin, A. (2016a). Improved transients in multiple frequencies estimation via dynamic regressor extension and mixing. *IFAC-PapersOnLine*, *49*(13), 99-104. doi: 10.1016/j.ifacol.2016.07.934

Aström, K., & Eykhoff, P. (1971). System identification—a survey. *Automatica*, *7*(2), 123 - 162. doi: 10.1016/0005-1098(71)90059-8

Åström, K., & Wittenmark, B. (2013). *Adaptive control*. Mineola, New York: Dover Publications.

Bartle, R. G., & Sherbert, D. R. (2000). *Introduction to real analysis*. New York: Wiley.

Bateman, H., & Erdélyi, A. (1953). *Higher transcendental functions II*. New York: McGraw-Hill.

Bazylev, D. N., Doria-Cerezo, A., Pyrkin, A., Bobtsov, A., & Ortega, R. (2017). A new approach for flux and rotor resistance estimation of induction motors. *IFAC-PapersOnLine*, *50*(1), 1885-1890. doi: 10.1016/j.ifacol.2017.08.259

Belkoura, L., Floquet, T., Ibn Taarit, K., Perruquetti, W., & Orlov, Y. (2011). Estimation problems for a class of impulsive systems. *Int. J. Robust Nonlinear Control*, *21*(10), 1066-1079. doi: 10.1002/rnc.1621

Belov, A., Ortega, R., & Bobtsov, A. (2018). Parameter identification of linear discrete-time systems with guaranteed transient performance. *IFAC-PapersOnLine*, *51*(15), 1038-1043. doi: 10.1016/j.ifacol.2018.09.057

Betta, G., Liguori, C., Paolillo, A., & Pietrosanto, A. (2002). A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis. *IEEE Trans. Instrum. Meas.*, *51*(6), 1316-1322. doi: 10.1109/TIM.2002.807987

Blackman, R. B. (1965). *Linear data-smoothing and prediction in theory and practice*. Reading: Addison-Wesley.

Bobtsov, A., Bazylev, D., Pyrkin, A., Aranovskiy, S., & Ortega, R. (2017). A robust nonlinear position observer for synchronous motors with relaxed excitation conditions. *Int. J. Control*, *90*(4), 813-824. doi: 10.1080/00207179.2016.1230229

Bobtsov, A. A., Pyrkin, A. A., Ortega, R., & Vedyakov, A. A. (2018). A state observer for sensorless control of magnetic levitation systems. *Automatica*, *97*, 263-270. doi: 10.1016/j.automatica.2018.08.004

Bourgeot, J., & Delaleau, E. (2007). Fast algebraic impact times estimation for a linear system subject to unilateral constraint. In *Proc. of the 46th IEEE Conf. on Decision and Control* (p. 2701-2706). New Orleans, LA, USA. doi: 10.1109/CDC.2007.4434023

Braci, M., & Diop, S. (2003). On numerical differentiation algorithms for nonlinear estimation. In *Proc. of the 42nd IEEE Conf. on Decision and Control* (Vol. 3, p. 2896-2901). Maui, HI, USA. doi: 10.1109/CDC.2003.1273065

Bromba, M. U. A., & Ziegler, H. (1983). On Hilbert space design of least-weighted-squares digital filters. *Int. J. Circ. Theor. Appl.*, *11*(1), 7-32. doi: 10.1002/cta.4490110103

Burch, N., Fishback, P. E., & Gordon, R. (2005). The least-squares property of the Lanczos derivative. *Math. Mag.*, *78*(5), 368-378. doi: 10.2307/30044192

Byrd, R. H., Hribar, M. E., & Nocedal, J. (1999). An interior point algorithm for large-scale nonlinear programming. *SIAM J. Optim.*, *9*(4), 877-900. doi: 10.1137/S1052623497325107

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, *8*(6), 679-698. doi: 10.1109/TPAMI.1986.4767851

Cerrada, M., Sánchez, R.-V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J., & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.*, *99*, 169-196. doi: 10.1016/j.ymssp.2017.06.012

Chen, B., Li, P., Pin, G., & Parisini, T. (2016). Estimation of multi-sinusoidal signals: A

deadbeat methodology. In *Proc. of the IEEE 55th Conf. on Decision and Control* (p. 3763-3768). doi: 10.1109/CDC.2016.7798836

Chen, B., Pin, G., & Parisini, T. (2014). An adaptive observer-based estimator for multi-sinusoidal signals. In *Proc. of the 2014 American Control Conf.* (p. 3450-3455). doi: 10.1109/ACC.2014.6858738

Chen, C. K., & Lee, J.-H. (1995). Design of high-order digital differentiators using L_1 error criteria. *IEEE Trans. Circuits Syst. II. Analog Digit. Signal Process.*, *42*(4), 287-291. doi: 10.1109/82.378044

Chitour, Y. (2002). Time-varying high-gain observers for numerical differentiation. *IEEE Trans. Autom. Control*, *47*(9), 1565-1569. doi: 10.1109/TAC.2002.802740

Cioranescu, N. (1938). La généralisation de la première formule de la moyenne. *Enseign. Math.*, *37*, 292-302.

Courant, R., & Hilbert, D. (1924). *Methoden der mathematischen Physik 1*. Berlin: Springer.

Cullum, J. (1971). Numerical differentiation and regularization. *SIAM J. Numer. Anal.*, *8*(2), 254-265. doi: 10.1137/0708026

CWRU. (2015). *Case western reserve university bearing data center bearing data set.* `https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website`. (Accessed: 2021-03-11)

Dabroom, A. M., & Khalil, H. K. (1997). Numerical differentiation using high-gain observers. In *Proc. of the 36th IEEE Conf. on Decision and Control* (Vol. 5, p. 4790-4795). San Diego, CA, USA. doi: 10.1109/CDC.1997.649776

Dabroom, A. M., & Khalil, H. K. (1999). Discrete-time implementation of high-gain observers for numerical differentiation. *Int. J. Control*, *72*(17), 1523-1537. doi: 10.1080/002071799220029

Daga, A. P., Fasana, A., Marchesiello, S., & Garibaldi, L. (2009). *The Politechnico di Torino rolling bearing test rig.* `ftp://ftp.polito.it/people/DIRG_BearingData/`. (Accessed: 2021-03-16)

Daga, A. P., Fasana, A., Marchesiello, S., & Garibaldi, L. (2019). The Politecnico di torino rolling bearing test rig: Description and analysis of open access data. *Mech. Syst. Signal Process.*, *120*, 252-273.

D'Amico, M., & Ferrigno, G. (1992). Comparison between the more recent techniques for smoothing and derivative assessment in biomechanics. *Med. Biol. Eng. Comput.*, *30*(2), 193-204. doi: 10.1007/bf02446130

Davis, P., & Hersh, R. (1981). *The mathematical experience*. Boston: Houghton Mifflin.

de Azevedo, H. D. M., Araújo, A. M., & Bouchonneau, N. (2016). A review of wind turbine bearing condition monitoring: State of the art and challenges. *Renew. Sust. Energ. Rev.*, *56*, 368-379. doi: 10.1016/j.rser.2015.11.032

De Persis, C., & Isidori, A. (2001). A geometric approach to nonlinear fault detection and isolation. *IEEE Trans. Autom. Control*, *46*(6), 853-865. doi: 10.1109/9.928586

De Persis, C., & Isidori, A. (2002). On the design of fault detection filters with game-theoretic-optimal sensitivity. *Int. J. Robust Nonlinear Control*, *12*(8), 729-747. doi: 10.1002/rnc.713

Diekema, E., & Koornwinder, T. H. (2012). Differentiation by integration using orthogonal polynomials, a survey. *J. Approx. Theory*, *164*(5), 637-667. doi:

10.1016/j.jat.2012.01.003

Diop, S., & Fliess, M. (1991a). Nonlinear observability, identifiability, and persistent trajectories. In *Proc. of the 30th IEEE Conf. on Decision and Control* (Vol. 1, p. 714-719 vol.1). Brighton, UK. doi: 10.1109/CDC.1991.261405

Diop, S., & Fliess, M. (1991b). On nonlinear observability. In *Proc. of the 1991 European Control Conf.* (p. 152-157). Grenoble, France.

Diop, S., Fromion, V., & Grizzle, J. W. (2001). A resettable kalman filter based on numerical differentiation. In *Prof. of the 2001 European Control Conf.* (p. 1239-1244). Porto, Portugal. doi: 10.23919/ECC.2001.7076086

Diop, S., Grizzle, J., Moraal, P., & Stefanopoulou, A. (1994). Interpolation and numerical differentiation for observer design. In *Proc. of the 1994 American Control Conf.* (Vol. 2, p. 1329-1333). Baltimore, MD, USA. doi: 10.1109/ACC.1994.752275

Diop, S., Grizzle, J. W., & Chaplais, F. (2000). On numerical differentiation algorithms for nonlinear estimation. In *Proc. of the 39th IEEE Conf. on Decision and Control* (Vol. 2, p. 1133-1138). Sydney, NSW, Australia. doi: 10.1109/CDC.2003.1273065

Diop, S., & Martínez-Guerra, R. (2001). An algebraic and data derivative information approach to nonlinear system diagnosis. In *Proc. of the 2001 European Control Conf.* (p. 2334-2339). Porto, Portugal. doi: 10.23919/ECC.2001.7076274

Doetsch, G. (1974). *Introduction to the theory and application of the Laplace transformation*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-65690-3

Egan, W. F. (2003). *Practical RF system design*. Wiley-IEEE. doi: 10.1002/0471654094

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems*. Dordrecht: Kluwer Academic Publishers.

Erdélyi, A., Magnus, W., Oberhettinger, F., & Tricomi, F. G. (1954a). *Tables of integral transforms* (A. Erdélyi, Ed.) (No. Vol. II). New York, Toronto, London: McGraw-Hill Book Company, Inc., New York-Toronto-London.

Erdélyi, A., Magnus, W., Oberhettinger, F., & Tricomi, F. G. (1954b). *Tables of integral transforms* (Vol. I). New York, Toronto, London: McGraw-Hill Book Company Inc.

Esfandiari, F., & Khalil, H. K. (1992). Output feedback stabilization of fully linearizable systems. *Int. J. Control*, *56*(5), 1007-1037. doi: 10.1080/00207179208934355

Fedele, G., & Coluccio, L. (2010). A recursive scheme for frequency estimation using the modulating functions method. *Applied Mathematics and Computation*, *216*(5), 1393-1400. doi: 10.1016/j.amc.2010.02.039

Fliess, M. (2006). Analyse non standard du bruit. *C. R. Acad. Sci. Paris, Ser. I*, *342*(10), 797-802. doi: 10.1016/j.crma.2006.02.037

Fliess, M. (2008). Critique du rapport signal à bruit en communications numériques. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, *9*, 419-429.

Fliess, M., & Join, C. (2003). An algebraic approach to fault diagnosis for linear systems. In *Proc. CESA 2003, Lille, France.*

Fliess, M., Join, C., & Mboup, M. (2010). Algebraic change-point detection. *Appl. Algebra Eng. Commun. Comput.*, *21*(2), 131-143. doi: 10.1007/s00200-010-0119-z

Fliess, M., Join, C., Mboup, M., & Sira-Ramírez, H. (2004). Compression différentielle de transitoires bruités. *Comptes Rendus Mathématique*, *339*(11), 821-826. doi: 10.1016/j.crma.2004.10.003

Fliess, M., Join, C., & Mounier, H. (2004). An introduction to nonlinear fault diagnosis

with an application to a congested internet router. In S. Tarbouriech, C. T. Abdallah, & J. Chiasson (Eds.), *Advances in Communication Control Networks* (p. 327-343). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-31597-1_16

Fliess, M., Join, C., & Sira-Ramírez, H. (2004). Robust residual generation for linear fault diagnosis: an algebraic setting with examples. *Int. J. Control*, *77*(14), 1223-1242. doi: 10.1080/00207170420002437 4

Fliess, M., Join, C., & Sira-Ramírez, H. (2008). Non-linear estimation is easy. *Int. J. Model. Identif. Control.*, *4*(1), 12-27. doi: 10.1504/IJMIC.2008.020996

Fliess, M., Mboup, M., Mounier, H., & Sira-Ramírez, H. (2003). Questioning some paradigms of signal processing via concrete examples. In H. Sira-Ramírez & G. Silva-Navarro (Eds.), *Algebraic Methods in Flatness, Signal Processing and State Estimation* (p. pp. 1-21). Editorial Lagares.

Fliess, M., & Sira-Ramírez, H. (2003a). An algebraic framework for linear identification. *ESAIM - Control Optim. Calc. Var.*, *9*, 151-168. doi: 10.1051/cocv:2003008

Fliess, M., & Sira-Ramírez, H. (2003b). State reconstructors: A possible alternative to asymptotic observers and Kalman filters. In *Prod. CESA 2003.* Lille, France.

Fliess, M., & Sira-Ramírez, H. (2004). Control via state estimations of some nonlinear systems. *IFAC Proc. Volumes*, *37*(13), 41-48. doi: 10.1016/S1474-6670(17)31198-9

Fliess, M., & Sira-Ramírez, H. (2008). Closed-loop parametric identification for continuous-time linear systems via new algebraic techniques. In W. L. Garnier H. (Ed.), *Identification of Continuous-time Models from Sampled Data* (p. 363-391). London: Springer. doi: 10.1007/978-1-84800-161-9_13

Fourier, J. B. J. (1822). *Théorie analytique de la chaleur*. Paris: Imprimerie de Firmin Didot.

Fourier, J. B. J., & Freeman, A. (1878). *The analytical theory of heat*. Cambridge: The University Press.

Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *Automatica*, *26*(3), 459 - 474. doi: 10.1016/0005-1098(90)90018-D

Frank, P. M. (1992). Principles of model-based fault detection. *IFAC Proc. Volumes*, *25*(10), 213 - 220. doi: 10.1016/S1474-6670(17)50824-1

Franklin, G. F., Powell, J. D., Workman, M. L., et al. (1998). *Digital control of dynamic systems*. Menlo Park: Addison Wesley Longman Inc.

Garnier, H., & Wang, L. (Eds.). (2008). *Identification of continuous-time models from sampled data*. Springer London. doi: 10.1007/978-1-84800-161-9

Gehring, N. (2015). *Algebraische Methoden zur Parameteridentifikation für lineare unendlichdimensionale Systeme* (Doctoral dissertation, Saarland University, Germany). doi: 10.22028/d291-23080

Gehring, N., Rudolph, J., & Stauch, C. (2012a). Algebraic identification of fluid parameters using transmission line dynamics. In *Proc. of the 13th Mechatronics Forum International Conf.* (p. 930-935). Linz, Austria,.

Gehring, N., Rudolph, J., & Stauch, C. (2012b). Identification of transmission line parameters using algebraic methods. *Proc. Appl. Math. Mech.*, *12*(1), 727-728. doi: 10.1002/pamm.201210352

Gehring, N., Stauch, C., & Rudolph, J. (2016). Parameter identification, fault detection and localization for an electrical transmission line. In *Proc. of the 2016 European Control Conf.*

(p. 2090-2095). doi: 10.1109/ECC.2016.7810600

Gensior, A. (2009). *Beiträge zur flachheitsbasierten Regelung leistungselektronischer Systeme*. Düren, Germany: Shaker Verlag GmbH.

Gensior, A., Weber, J., Güldner, H., & Rudolph, J. (2007). An algebraic parameter identification algorithm and asymptotic observers for estimation of the load of a boost converter. In *Proc. of the 2007 IEEE International Symposium on Industrial Electronics* (p. 7-11). Vigo, Spain. doi: 10.1109/ISIE.2007.4374563

Gensior, A., Weber, J., Rudolph, J., & Güldner, H. (2008). Algebraic parameter identification and asymptotic estimation of the load of a boost converter. *IEEE Trans. Ind. Electron.*, *55*(9), 3352-3360. doi: 10.1109/TIE.2008.928102

Gertler, J. (1991). Analytical redundancy methods in fault detection and isolation - survey and synthesis. *IFAC Proceedings Volumes*, *24*(6), 9-21. doi: 10.1016/S1474-6670(17)51119-2

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. Baltimore: Johns Hopkins University Press.

Hanke, M., & Scherzer, O. (2001). Inverse problems light: Numerical differentiation. *Am. Math. Mon.*, *108*(6), 512. doi: 10.2307/2695705

Hao, D. N. (1994). A mollification method for ill-posed problems. *Numerische Mathematik*, *68*(4), 469-506. doi: 10.1007/s002110050073

Hegland, M., & Anderssen, R. S. (1998). A mollification framework for improperly posed problems. *Numerische Mathematik*, *78*(4), 549-575. doi: 10.1007/s002110050325

Hofmann, L. G., Lion, P. M., & Best, J. J. (1966). *Theoretical and experimental research on parameter tracking systems* (Final report NASA-CR-452). NASA.

Hou, M. (2012). Parameter identification of sinusoids. *IEEE Trans. Automat. Contr.*, *57*(2), 467-472. doi: 10.1109/TAC.2011.2164736

Hwang, I., Kim, S., Kim, Y., & Seah, C. E. (2010). A survey of fault detection, isolation, and reconfiguration methods. *IEEE Trans. Control Syst. Technol.*, *18*(3), 636-653. doi: 10.1109/TCST.2009.2026285

Isermann, R. (1977). *Digitale Regelsysteme*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-96341-4

Isermann, R. (2006). *Fault-diagnosis systems*. Berlin, Heidelberg: Springer. doi: 10.1007/3-540-30368-5

Isermann, R. (2011). *Fault-diagnosis applications*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-12767-0

Isermann, R. (2013a). *Identifikation dynamischer Systeme 1: Grundlegende Methoden*. Berlin, Heidelberg: Springer.

Isermann, R. (2013b). *Identifikation dynamischer Systeme 2: Besondere Methoden, Anwendungen*. Berlin, Heidelberg: Springer.

Isermann, R., & Ballé, P. (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Eng. Pract.*, *5*(5), 709 - 719. doi: 10.1016/S0967-0661(97)00053-1

Jenkins, B., Annaswamy, A. M., & Kojic, A. (2015). Matrix regressor adaptive observers for battery management systems. In *2015 IEEE International Symposium on Intelligent Control* (p. 707-714). Sydney, NSW, Australia. doi: 10.1109/ISIC.2015.7307293

Jiang, B., Staroswiecki, M., & Cocquempot, V. (2006). Fault accommodation for

nonlinear dynamic systems. *IEEE Trans. Autom. Control*, *51*(9), 1578-1583. doi: 10.1109/TAC.2006.878732

Join, C. (2002). *Diagnostic des systèmes non linéaires - Contribution aux méthodes de découplage* (Doctoral dissertation). Université Henri Poincaré - Nancy I, France.

Jouffroy, J., & Reger, J. (2015). Finite-time simultaneous parameter and state estimation using modulating functions. In *Proc. of the 2015 IEEE Conf. on Control Applications* (p. 394-399). Sydney, NSW, Australia.

Kaczmarz, S., & Steinhaus, H. (1951). *Theorie der Orthogonalreihen*. New York: Chelsea Publishing Company.

Kailath, T. (1980). *Linear systems*. Englewood Cliffs: Prentice-Hall.

Kantz, H., & Schreiber, T. (2004). *Nonlinear time series analysis*. Cambridge: Cambridge University Press.

Khalil, H. K. (2000). Universal integral controllers for minimum-phase nonlinear systems. *IEEE Trans. on Autom. Control*, *45*(3), 490-494. doi: 10.1109/9.847730

Khalil, H. K. (2002). *Nonlinear systems*. Upper Saddle River: Prentice Hall.

Khalil, H. K. (2009). Analysis of sampled-data high-gain observers in the presence of measurement noise. *Eur. J. Control*, *15*(2), 166 - 176. doi: 10.3166/ejc.15.166-176

Khalil, H. K. (2017). *High-gain observers in nonlinear feedback control*. Philadelphia: SIAM. doi: 10.1137/1.9781611974867

Kiltz, L. (2017). *Algebraische Ableitungsschätzer in Theorie und Anwendung* (Doctoral dissertation, Saarland University, Germany). doi: 10.22028/D291-27034

Kiltz, L., Braun, T., Reuß, N., Lüttgens, A., Bode, C., König, O., ... Ettner, K. (2021). *Sensorlose Überwachung* (No. DE 10 2019 219 189.4). Deutsches Patent- und Markenamt.

Kiltz, L., Janocha, M., & Rudolph, J. (2013). Algebraic estimation of impact times: Juggling a ball with a magnetically levitated plate. In *Proc. of the 2nd International Conf. on Systems and Computer Science* (p. 145-149). Villeneuve d'Ascq, France. doi: 10.1109/IcConSCS.2013.6632038

Kiltz, L., Join, C., Mboup, M., & Rudolph, J. (2014). Fault-tolerant control based on algebraic derivative estimation applied on a magnetically supported plate. *Control Eng. Pract.*, *26*, 107 - 115. doi: 10.1016/j.conengprac.2014.01.009

Kiltz, L., Mboup, M., & Rudolph, J. (2012). Fault diagnosis on a magnetically supported plate. In *Proc. of the 1st International Conf. on Systems and Computer Science.* Lille, France. doi: 10.1109/IConSCS.2012.6502453

Kiltz, L., & Rudolph, J. (2013). Parametrization of algebraic numerical differentiators to achieve desired filter characteristics. In *Proc. of the 52nd IEEE Conf. on Decision and Control* (p. 7010-7015). Florence, Italy. doi: 10.1109/CDC.2013.6761000

Kiltz, L., Rueß, N., Braun, T., Lüttgens, A., Bode, C., König, O., ... Ettner, K. (2019). ZF Rotatronic shifter—Robust noise reduction via real-time embedded learning algorithms. CTI Symposium, Berlin, Germany.

Kiss, I. Z., Lv, Q., & Hudson, J. L. (2005). Synchronization of non-phase-coherent chaotic electrochemical oscillations. *Phys. Rev. E*, *71*, 035201. doi: 10.1103/PhysRevE.71.035201

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, *50*(5), 363-370. doi: 10.1007/BF00336961

Kotelnikov, V. A. (1933). On the transmission capacity of the 'ether' and of cables in electrical communications. In *Proc. 1st All-Union Conf. on the technological reconstruction of the communications sector and the development of low-current engineering* (p. 1-23). Moscow, UdSSR.

Kreisselmeier, G. (1977). Adaptive observers with exponential rate of convergence. *IEEE Trans. Autom. Control*, *22*(1), 2-8. doi: 10.1109/TAC.1977.1101401

Lanczos, C. (1956). *Applied analysis*. Englewood Cliffs: Prentice Hall, Inc.

Lebedev, N. N., & Silverman, R. A. (1965). *Special functions and their applications*. Englewood Cliffs: Prentice-Hall.

Levant, A. (1998). Robust exact differentiation via sliding mode technique. *Automatica*, *34*(3), 379 - 384. doi: 10.1016/S0005-1098(97)00209-4

Levant, A. (2003). Higher-order sliding modes, differentiation and output-feedback control. *Int. J. Control*, *76*(9-10), 924-941. doi: 10.1080/0020717031000099029

Li, C., Sanchez, V., Zurita, G., Cerrada Lozada, M., & Cabrera, D. (2016). Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement. *ISA Trans.*, *60*, 274-284. doi: 10.1016/j.isatra.2015.10.014

Li, P. (2019). *Finite-time system identification, estimation and fault detection* (Doctoral dissertation, Imperial College London, UK). doi: 10.25560/70626

Li, P., Boem, F., Pin, G., & Parisini, T. (2018). Deadbeat simultaneous parameter-state estimation for linear continuous-time systems: a kernel-based approach. In *Proc. of the 2018 European Control Conf.* (p. 2493-2498). Limassol, Cyprus.

Li, P., Boem, F., Pin, G., & Parisini, T. (2020). Kernel-based simultaneous parameter-state estimation for continuous-time systems. *IEEE Trans. Autom. Control*, *65*(7), 3053-3059. doi: 10.1109/TAC.2019.2953146

Lion, P. M. (1967). Rapid identification of linear and nonlinear systems. *AIAA Journal*, *5*(10), 1835-1842. doi: 10.2514/3.4313

Liu, D. Y. (2011). *Analyse d'erreurs d'estimateurs des dérivées de signaux bruités et applications* (Doctoral dissertation). Université des Sciences et Technologies de Lille - Lille 1, France.

Liu, D. Y., Gibaru, O., & Perruquetti, W. (2010). Convergence rate of the causal Jacobi derivative estimator. In J.-D. Boissonnat et al. (Eds.), *Curves and surfaces* (p. 445-455). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-27413-8_28

Liu, D. Y., Gibaru, O., & Perruquetti, W. (2011a). Differentiation by integration with Jacobi polynomials. *J. Comput. Appl. Math.*, *235*(9), 3015-3032. doi: 10.1016/j.cam.2010.12.023

Liu, D. Y., Gibaru, O., & Perruquetti, W. (2011b). Error analysis of Jacobi derivative estimators for noisy signals. *Numer. Algorithms*, *58*(1), 53-83. doi: 10.1007/s11075-011-9447-8

Liu, D. Y., Gibaru, O., Perruquetti, W., Fliess, M., & Mboup, M. (2008). An error analysis in the algebraic estimation of a noisy sinusoidal signal. In *Proc. of the 16th Mediterranean Conf. on Control and Automation* (p. 1296-1301). Ajaccio, France. doi: 10.1109/MED.2008.4602161

Liu, D. Y., Laleg-Kirati, T. M., Perruquetti, W., & Gibaru, O. (2014). Non-asymptotic state estimation for a class of linear time-varying systems with unknown inputs. *IFAC Proceedings Volumes*, *47*(3), 3732-3738. doi: 10.3182/20140824-6-ZA-1003.02036

Ljung, L. (1999). *System identification: Theory for the user*. Upper Saddle River, New Jersey: Prentice Hall PTR.

Loeb, J., & Cahen, G. (1965). More about process identification. *IEEE Trans. Autom. Control*, *10*(3), 359-361. doi: 10.1109/TAC.1965.1098172

Lomakin, A., & Deutscher, J. (2020a). Algebraic fault detection and identification for rigid robots. In *Proc. of the 2020 IEEE International Conf. on Robotics and Automation* (p. 9352-9358). Paris, France. doi: 10.1109/ICRA40945.2020.9197561

Lomakin, A., & Deutscher, J. (2020b). Identification of dynamic parameters for rigid robots based on polynomial approximation. In *Proc. of the 2020 IEEE/RSJ International Conf. on Intelligent Robots and Systems* (p. 7271-7278). Las Vegas, NV, USA. doi: 10.1109/IROS45743.2020.9341195

Lomakin, A., & Deutscher, J. (2021). Algebraische Detektion und Identifikation von Parameterfehlern in der Robotik. *at - Automatisierungstechnik*, *69*(8), 645-655. doi: doi:10.1515/auto-2021-0045

Lomakin, A., & Deutscher, J. (2022). Reliable algebraic fault detection and identification of robots. *IEEE Transactions on Automation Science and Engineering*, 1-17. doi: 10.1109/TASE.2021.3137182

Luenberger, D. G. (1997). *Optimization by vector space methods*. New York, London, Sydney, Toronto: John Wiley & Sons.

Madden, H. (1978). Comments on the Savitzky-Golay convolution method for least-squares fit smoothing and differentiation of digital data. *Anal. Chem.*, *50*(9), 1383-86. doi: 10.1021/ac50031a048

Mai, P., & Hillermeier, C. (2008). New insights into derivative estimation via least squares approximation - theory and application. In *Proc. of the 2008 American Control Conf.* (p. 2427-2434). Seattle, WA, USA. doi: 10.1109/ACC.2008.4586855

Mai, P., & Hillermeier, C. (2010). Fault tolerant tracking control for nonlinear systems based on derivative estimation. In *Proc. of the 2010 American Control Conf.* (p. 6486-6493). Baltimore, MD, USA. doi: 10.1109/ACC.2010.5531443

Mai, P., & Hillermeier, C. (2010). Fehlertolerante Regelung bei aktuatorähnlichen Fehlern mittels Ableitungsschätzung. *at - Automatisierungstechnik*, *58*(2). doi: 10.1524/auto.2010.0820

Mallat, S. (1999). *A wavelet tour of signal processing*. Cambridge, Massachusetts, USA: Academic Press.

Marino, R., & Tomei, P. (2002). Global estimation of n unknown frequencies. *IEEE Trans. Automat. Contr.*, *47*(8), 1324-1328. doi: 10.1109/TAC.2002.800761

Marr, D., & Hildreth, E. C. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*, 187-217. doi: 10.1098/rspb.1980.0020

Mboup, M. (2009). Parameter estimation for signals described by differential equations. *Appl. Anal.*, *88*(1), 29-52. doi: 10.1080/00036810802555441

Mboup, M., Join, C., & Fliess, M. (2007). A revised look at numerical differentiation with an application to nonlinear feedback control. In *Proc. of the 15th Mediterranean Conf. on Control and Automation*. Athen, Greece. doi: 10.1109/MED.2007.4433728

Mboup, M., Join, C., & Fliess, M. (2009). Numerical differentiation with annihilators in noisy environment. *Numer. Algorithms*, *50*(4), 439-467. doi: 10.1007/s11075-008-9236-1

Mboup, M., Join, C., Fliess, M., Wang, Y., Zheng, G., Efimov, D., & Perruquetti, W. (2020).

Comments on 'differentiator application in altitude control for an indoor blimp robot'. *Int. J. Control*, *93*(5), 1218-1219.

Mboup, M., & Riachy, S. (2014). A frequency domain interpretation of the algebraic differentiators. *IFAC Proc. Volumes*, *47*(3), 9147-9151. doi: 10.3182/20140824-6-za-1003.02132

Mboup, M., & Riachy, S. (2018). Frequency-domain analysis and tuning of the algebraic differentiators. *Int. J. Control*, *91*(9), 2073-2081. doi: 10.1080/00207179.2017.1421776

McFadden, P. D., & Smith, J. D. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *J. Sound Vib.*, *96*(1), 69-82. doi: 10.1016/0022-460X(84)90595-9

McFadden, P. D., & Smith, J. D. (1985). The vibration produced by multiple point defects in a rolling element bearing. *J. Sound Vib.*, *98*(2), 263-273. doi: 10.1016/0022-460X(85)90390-6

Meinardus, G. (1967). *Approximation of functions: Theory and numerical methods*. Berlin, Heidelberg, Germany: Springer. doi: 10.1007/978-3-642-85643-3

Mikusinski, J. (1983). *Operational calculus*. Oxford, New York, Toronto, Sydney, Paris, Frankfurt: Pergamon Press.

Murio, D. A. (2011). *The mollification method and the numerical solution of ill-posed problems*. New York, Santa Barbara, London, Sydney, Toronto: John Wiley & Sons. doi: 10.1002/9781118033210

Narendra, K. S., & Annaswamy, A. M. (2012). *Stable adaptive systems*. New York: Dover Publications.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). PRONOSTIA : An experimental platform for bearings accelerated degradation tests. In *Proc. of the 2012 IEEE International Conf. on Prognostics and Health Management* (p. 1-8). Denver, Colorado, USA.

Nikolaou, N. G., & Antoniadis, I. A. (2002). Rolling element bearing fault diagnosis using wavelet packets. *NDT & E International*, *35*(3), 197 - 205. doi: 10.1016/S0963-8695(01)00044-5

Nöthen, C. (2007). *Beiträge zur Rekonstruktion nicht direkt gemessener Größen bei der Silizium-Einkristallzüchtung nach dem Czochralski-Verfahren* (Diplomarbeit). Technische Universität Dresden, Germany.

Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Trans. AIEE*, *47*(2), 617-644. doi: 10.1109/T-AIEE.1928.5055024

Oppenheim, A. V., & Schafer, R. W. (1975). *Digital signal processing*. Englewood Cliffs: Prentice-Hall.

Oppenheim, A. V., & Schafer, R. W. (1989). *Discrete-time signal processing*. Englewood Cliffs: Prentice Hall.

Ortega, R., Aranovskiy, S., Pyrkin, A., Astolfi, A., & Bobtsov, A. (2021). New results on parameter estimation via dynamic regressor extension and mixing: Continuous and discrete-time cases. *IEEE Trans. Automat. Contr.*, *66*(5), 2265-2272. doi: 10.1109/TAC.2020.3003651

Ortega, R., Gerasimov, D. N., Barabanov, N. E., & Nikiforov, V. O. (2019). Adaptive control of linear multivariable systems using dynamic regressor extension and mixing estimators: Removing the high-frequency gain assumptions. *Automatica*, *110*, 108589.

doi: 10.1016/j.automatica.2019.108589

Ortega, R., Nikiforov, V., & Gerasimov, D. (2020). On modified parameter estimators for identification and adaptive control. A unified framework and some new schemes. *Annual Reviews in Control*, *50*, 278-293. doi: 10.1016/j.arcontrol.2020.06.002

Ortega, R., Praly, L., Aranovskiy, S., Yi, B., & Zhang, W. (2018). On dynamic regressor extension and mixing parameter estimators: Two Luenberger observers interpretations. *Automatica*, *95*, 548-551. doi: 10.1016/j.automatica.2018.06.011

Othmane, A. (2021). *AlgDiff: A Python package with MATLAB coupling implementing all necessary tools for the design, analysis, and discretization of algebraic differentiators.* (Available at `https://github.com/aothmane-control/Algebraic-differentiators`, version 1.0.0) doi: 10.5281/zenodo.5172198

Othmane, A., Kiltz, L., & Rudolph, J. (2022). Survey on algebraic numerical differentiation: Historical developments, parametrization, examples, and applications. *Int. J. Syst. Sci.*, *53*(9), 1848-1887. (free access) doi: 10.1080/00207721.2022.2025948

Othmane, A., Mounier, H., & Rudolph, J. (2021). Parametrization of algebraic differentiators for disturbance annihilation with an application to the differentiation of quantized signals. *IFAC-PapersOnLine*, *54*(9), 335-340. doi: 10.1016/j.ifacol.2021.06.091

Othmane, A., & Rudolph, J. (2021). Data and computation efficient model-based fault detection for rolling element bearings using numerical differentiation. In *Proc. of the 5th International Conf. on Control and Fault-Tolerant Systems* (p. 163-168). Saint-Raphaël, France: IEEE. doi: 10.1109/SysTol52990.2021.9595974

Othmane, A., Rudolph, J., & Mounier, H. (2020). Analysis of the parameter estimate error when algebraic differentiators are used in the presence of disturbances. *IFAC-PapersOnLine*, *53*(2), 572-577. doi: 10.1016/j.ifacol.2020.12.493

Othmane, A., Rudolph, J., & Mounier, H. (2021a). Systematic comparison of numerical differentiators and an application to model-free control. *Eur. J. Control*, *62*, 113-119. doi: 10.1016/j.ejcon.2021.06.020

Othmane, A., Rudolph, J., & Mounier, H. (2021b). Systematic comparison of numerical differentiators and an application to model-free control. *Eur. J. Control*, *62*, 113-119. doi: 10.1016/j.ejcon.2021.06.020

Papoulis, A. (1962). *The Fourier integral and its applications*. New York, San Francisco, London, Toronto: McGraw-Hill.

Papoulis, A. (1984). *Probability, random variables, and stochastic processes*. New York: McGraw-Hill.

Patton, R. J., Frank, P. M., & Clark, R. N. (Eds.). (1989). *Fault diagnosis in dynamic systems: Theory and application*. Upper Saddle River: Prentice-Hall, Inc.

Pearson, A., & Lee, F. (1985). On the identification of polynomial input-output differential systems. *IEEE Trans. Automat. Control*, *30*(8), 778-782. doi: 10.1109/TAC.1985.1104051

Persson, P.-O., & Strang, G. (2003). Smoothing by Savitzky-Golay and Legendre filters. In J. Rosenthal & D. S. Gilliam (Eds.), *Proc. of mathematical Systems Theory in Biology, Communications, Computation, and Finance* (p. 301-315). New York, NY: Springer.

Peter, K., & Isermann, R. (1990). Parameter-adaptive PID-control based on continuous-time process models. *IFAC Proc. Volumes*, *23*(1), 241-246. doi: 10.1016/S1474-6670(17)52727-5

Pettit, J. M., & McWhorter, M. M. (1961). *Electronic amplifier circuits, theory and design*. New York: McGraw-Hill.

Pin, G., Assalone, A., Lovera, M., & Parisini, T. (2016). Non-asymptotic kernel-based parametric estimation of continuous-time linear systems. *IEEE Trans. Automat. Contr.*, *61*(2), 360-373. doi: 10.1109/TAC.2015.2434075

Pin, G., Lovera, M., Assalone, A., & Parisini, T. (2013). Kernel-based non-asymptotic state estimation for linear continuous-time systems. In *Proc. of the 2013 American Control Conf.* (p. 3123-3128). Washington, DC, USA. doi: 10.1109/ACC.2013.6580311

Preisig, H. A., & Rippin, D. W. T. (1993a). Theory and application of the modulating function method— I. Review and theory of the method and theory of the spline-type modulating functions. *Comput. Chem. Eng.*, *17*(1), 1-16. doi: 10.1016/0098-1354(93)80001-4

Preisig, H. A., & Rippin, D. W. T. (1993b). Theory and application of the modulating function method—II. Algebraic representation of Maletinsky's spline-type modulating functions. *Comput. Chem. Eng.*, *17*(1), 17 - 28. doi: 10.1016/0098-1354(93)80002-5

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). New York: Cambridge University Press.

Qian, L. (2019). *Observer-based fault detection and estimation of rolling element bearing systems*. Düren: Shaker Verlag GmbH. doi: 10.2370/9783844066982

Rader, C. M., & Jackson, L. B. (2006). Approximating noncausal IIR digital filters having arbitrary poles, including new Hilbert transformer designs, via forward/backward block recursion. *IEEE Trans. on Circuits and Systems I: Regular Papers*, *53*(12), 2779-2787. doi: 10.1109/TCSI.2006.883877

Raffel, M., Willert, C. E., & Kompenhans, J. (2007). *Particle image velocimetry: A practical guide*. Berlin, Heidelberg: Springer.

Rangarajan, S., & Purushothaman, S. P. (2005). Lanczos' generalized derivative for higher orders. *J. Comput. Appl. Math.*, *177*(2), 461-465. doi: 10.1016/j.cam.2004.10.016

Reger, J., & Jouffroy, J. (2009). On algebraic time-derivative estimation and deadbeat state reconstruction. In *Proc. of the 48h IEEE Conf. on Decision and Control held jointly with 2009 28th Chinese Control Conf.* (p. 1740-1745). Shanghai, China.

Reger, J., Sira-Ramírez, H., & Fliess, M. (2005). On non-asymptotic observation of nonlinear systems. In *Proc. of the 44th IEEE Conf. on Decision and Control* (p. 4219-4224). Seville, Spain.

Riachy, S., Efimov, D., & Mboup, M. (2016). Universal integral control: An approach based on mollifiers. *IEEE Trans. Autom. Control*, *61*(1), 204-209. doi: 10.1109/TAC.2015.2427631

Sastry, S., & Bodson, M. (1989). *Adaptive control: Stability, convergence, and robustness*. Upper Saddle River: Prentice-Hall, Inc.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, *36*(8), 1627-1639. doi: 10.1021/ac60214a047

Schäfer, C., Rosenblum, M. G., Abel, H.-H., & Kurths, J. (1999). Synchronization in the human cardiorespiratory system. *Phys. Rev. E*, *60*, 857-870. doi: 10.1103/PhysRevE.60.857

Schwartz, L. (1966). *Théorie Des Distributions.* Paris: Hermann.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, *27*(3), 379-423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1949). Communication in the presence of noise. *Proc. of the IRE*, *37*(1), 10-21. doi: 10.1109/JRPROC.1949.232969

Shinbrot, M. (1954). *On the analysis of linear and nonlinear dynamical systems from transient-response data* (Tech. Rep.). Ames Aeronautical Laboratory, Moffett Field, Calif. (NACA TN 3288)

Shinbrot, M. (1957). On the analysis of linear and nonlinear systems. *Trans. ASME*, *79*(3), 547-552. doi: 10.1115/1.4013092

Sidhom, L. (2011). *Sur les différentiateurs en temps réel: algorithmes et applications* (Doctoral dissertation). INSA de Lyon, France.

Slotine, J., & Li, W. (1991). *Applied nonlinear control*. Englewood Cliffs, New Jersey: Prentice Hall.

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mech. Syst. Signal Process.*, *64-65*, 100-131. doi: 10.1016/j.ymssp.2015.04.021

Stack, J. R., Habetler, T. G., & Harley, R. G. (2004). Fault classification and fault signature production for rolling element bearings in electric machines. *IEEE Trans. Ind. Appl.*, *40*(3), 735-739. doi: 10.1109/TIA.2004.827454

Stauch, C., Gehring, N., & Rudolph, J. (2013). Algebraic parameter identification for infinite dimensional fluid transmission line models. *Proc. Inst. Mech. Eng., I: J. Syst. Control Eng.*, *227*(10), 733-743. doi: 10.1177/095965181 3498716

Stewart, G. W. (1998). On the adjugate matrix. *Linear Algebra Appl.*, *283*(1), 151-164. doi: 10.1016/S0024-3795(98)10098-8

Stroud, A. H. (1974). *Numerical quadrature and solution of ordinary differential equations*. New York: Springer. doi: 10.1007/978-1-4612-6390-6

Szegö, G. (1939). *Orthogonal polynomials*. New York: AMS.

Tietze, U., Schenk, C., & Gamm, E. (2008). *Electronic Circuits*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-78655-9

Tiganj, Z., & Mboup, M. (2009). Spike detection and sorting: Combining algebraic differentiations with ICA. In T. Adali, C. Jutten, J. M. T. Romano, & A. K. Barros (Eds.), *Independent Component Analysis and Signal Separation* (p. 475-482). Berlin, Heidelberg: Springer.

Tisserand, E., Lezama, J., Schweitzer, P., & Berviller, Y. (2015). Series arcing detection by algebraic derivative of the current. *Electr. Power Syst. Res.*, *119*, 91-99. doi: 10.1016/j.epsr.2014.09.011

Trapero, J. R., Sira-Ramírez, H., & Batlle, V. F. (2007). An algebraic frequency estimator for a biased and noisy sinusoidal signal. *Signal Process.*, *87*(6), 1188-1201. doi: 10.1016/j.sigpro.2006.10.006

Trapero, J. R., Sira-Ramírez, H., & Batlle, V. F. (2008). On the algebraic identification of the frequencies, amplitudes and phases of two sinusoidal signals from their noisy sum. *Int. J. Control*, *81*(3), 507-518. doi: 10.1080/00207170701561419

Tricomi, F. G. (1970). *Vorlesungen über Orthogonalreihen*. Berlin, Heidelberg: Springer.

Turin, G. L. (1960). An introduction to matched filters. *IRE Trans. Inf. Theory*, *6*(3), 311-329. doi: 10.1109/TIT.1960.1057571

Unbehauen, H., & Rao, G. P. (1987). *Identification of continuous systems*. New York: Elsevier Science Inc.

Ushirobira, R. (2018). Algebraic differentiators through orthogonal polynomials series expansions. *Int. J. Control*, *91*(9), 2082-2089. doi: 10.1080/00207179.2017.1406151

Ushirobira, R., Perruquetti, W., & Mboup, M. (2016). An algebraic continuous time parameter estimation for a sum of sinusoidal waveform signals. *Int. J. Adapt. Control Signal Process.*, *30*(12), 1689-1713. doi: 10.1002/acs.2688

Ushirobira, R., Perruquetti, W., Mboup, M., & Fliess, M. (2012). Algebraic parameter estimation of a biased sinusoidal waveform signal from noisy data. *IFAC Proceedings Volumes*, *45*(16), 167-172. doi: 10.3182/20120711-3-BE-2027.00393

Ushirobira, R., Perruquetti, W., Mboup, M., & Fliess, M. (2013). Algebraic parameter estimation of a multi-sinusoidal waveform signal from noisy data. In *Proc. of the 2013 European Control Conf.* (p. 1902-1907). Zurich, Switzerland. doi: 10.23919/ECC.2013.6669742

Ushirobira, R., & Quadrat, A. (2016). Algebraic estimation of a biased and noisy continuous signal via orthogonal polynomials. In *55th ieee conf. on decision and control* (p. 359-364). doi: 10.1109/CDC.2016.7798295

Vasiljevic, L. K., & Khalil, H. K. (2006). Differentiation with high-gain observers the presence of measurement noise. In *Proc. of the 45th IEEE Conf. on Decision and Control* (p. 4717-4722). San Diego, CA, USA. doi: 10.1109/CDC.2006.377230

Vasiljevic, L. K., & Khalil, H. K. (2008). Error bounds in differentiation of noisy signals by high-gain observers. *Syst. Control. Lett.*, *57*(10), 856 - 862. doi: 10.1016/j.sysconle.2008.03.018

Wang, Y., Zheng, G., Efimov, D., & Perruquetti, W. (2018). Differentiator application in altitude control for an indoor blimp robot. *Int. J. Control*, *91*(9), 2121-2130. doi: 10.1080/00207179.2018.1441549

Watson, G. N. (1995). *A treatise on the theory of Bessel functions*. Cambridge: Cambridge University Press.

Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, *12*(6), 601 - 611. doi: 10.1016/0005-1098(76)90041-8

Witkin, A. P. (1987). Scale-space filtering. In M. A. Fischler & O. Firschein (Eds.), *Readings in computer vision* (p. 329-332). San Francisco (CA): Morgan Kaufmann. doi: 10.1016/b978-0-08-051581-6.50036-2

Xia, X. (2002). Global frequency estimation using adaptive identifiers. *IEEE Trans. Automat. Contr.*, *47*(7), 1188-1193. doi: 10.1109/TAC.2002.800670

Yan, X., Primot, M., & Plestan, F. (2014). Comparison of differentiation schemes for the velocity and acceleration estimations of a pneumatic system. *IFAC Proc. Volumes*, *47*(3), 49-54. doi: 10.3182/20140824-6-ZA-1003.01963

Yan, X.-G., & Edwards, C. (2007). Nonlinear robust fault reconstruction and estimation using a sliding mode observer. *Automatica*, *43*(9), 1605 - 1614. doi: 10.1016/j.automatica.2007.02.008

Yang, H., & Saif, M. (1996). Monitoring and diagnostics of a class of nonlinear systems using a nonlinear unknown input observer. In *Proc. of the 1996 IEEE International Conf. on control applications held together with IEEE International symposium on intelligent control* (p. 1006-1011). Dearborn, MI, USA. doi: 10.1109/CCA.1996.559053

Yi, B., & Ortega, R. (2022). Conditions for convergence of dynamic regressor extension and mixing parameter estimators using LTI filters. *IEEE Trans. Automat. Contr.*. doi:

10.1109/TAC.2022.3149964

Yi, B., Ortega, R., Siguerdidjane, H., & Zhang, W. (2018). An adaptive observer for sensorless control of the levitated ball using signal injection. In *Proc. of the 2018 IEEE Conf. on Decision and Control* (p. 6882-6887). Miami Beach, FL, USA. doi: 10.1109/CDC.2018.8619375

Young, P. (1981). Parameter estimation for continuous-time models - A survey. *Automatica*, *17*(1), 23-39. doi: 10.1016/0005-1098(81)90082-0

Yung, C., & Bonnett, A. H. (2004). Repair or replace? *IEEE Ind. Appl. Mag.*, *10*(5), 48-58. doi: 10.1109/MIA.2004.1330770

Zivanovic, R. (2007). An adaptive differentiation filter for tracking instantaneous frequency in power systems. *EEE Trans. Power Deliv.*, *22*(2), 765-771. doi: 10.1109/TPWRD.2007.893368

Othmane and Rudolph (2021)

Othmane, A., & Rudolph, J. Data and computation efficient model-based fault detection for rolling element bearings using numerical differentiation. In Proc. of the 5th International Conf. on Control and Fault-Tolerant Systems (p. 163-168). Saint-Raphaël, France: IEEE. doi: 10.1109/SysTol52990.2021.9595974 (2021).