

# Localized State Space and Polynomial Filters with Applications in Electrocardiography

A thesis submitted to attain the degree of  
Doctor of Sciences of ETH Zurich  
(Dr. sc. ETH Zurich)

presented by

**Reto A. Wildhaber**

Dr. med., University of Zurich  
born on September 10, 1979  
citizen of Unterägeri (ZG)

accepted on the recommendation of  
Prof. Dr. Hans-Andrea Loeliger, examiner  
Prof. Dr. Ali H. Sayed, co-examiner

2019

**Series in Signal and Information Processing**

**Vol. 32**

**Editor: Hans-Andrea Loeliger**

**Bibliographic Information published by Die Deutsche Nationalbibliothek**

Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data is available in the internet at <http://dnb.d-nb.de>.

Copyright © 2019 by Reto Wildhaber

First Edition 2019

HARTUNG-GORRE VERLAG KONSTANZ

ISSN 1616-671X

ISBN-10: 3-86628-652-X

ISBN-13: 978-3-86628-652-8

# Acknowledgments

After more than five years as a PhD student at ETH Zürich and after being involved in an amazing project at Bern University of Applied Sciences (BFH) in Biel for over six years, it's time to appreciate all the support I received over the past years.

I address my thanks in chronological order. It was during my studies in medicine, when I first met Rolf Vogel. Rolf offered me the opportunity to gather experiences in his research team at ARTORG in Bern. Today, more than 7 years later, it's once again thanks to Rolf and the financial support of the *Solothurner Spitäler* that I can continue my research.

Likewise, my sincere thanks go to Marcel Jacomet and Josef Goette. Marcel employed me and gave me financial guarantee to do a PhD at his institute at BFH, and provided me with the confidence and manpower to realize my ideas. Josef continuously supported me with his immense knowledge, his curiosity, and his endless patience. It was also Josef, who proofread *all* my thesis and gave me valuable feedback of finest quality.

My sincere thanks equally go to my supervisor and professor at ETH Zürich, Hans-Andrea Loeliger. He accepted me as a doctoral student in his group despite (or due to) my uncommon curriculum vitae, supported me with his enormous expertise, and gave me his strong support until the end of my thesis—and even beyond. I am also grateful to Ali H. Sayed for taking the time to read and examine my thesis.

Further, my thanks go to Nour Zalmai. He opened my eyes to what signal processing is capable of and always unreservedly shared his knowledge with me in endless discussions; I'm still weighing whether he was a workmate or rather a teacher to me—or both at once. Regardless, I'm thankful for the shared PhD time and the friendship that has grown over the past years.

My work was also highly influenced and pushed forward by fruitful discussion with Hampus Malmberg, who always found time to listen

when I was stuck with my thoughts; and equally Federico Wadehn (linear algebra and probability problems), Elisabeth Ren (for her valuable feedbacks and for sharing her ideas with me), and later Raphael Keusch, Boxiao Ma, and Patrick Murer. They all left their marks in my thesis.

Working with students of different backgrounds was always a pleasure and they often taught me new aspects I had not thought about before. I like to mention some exceptional students: Desirée Clausen (cardiac dipole model, master thesis ETH, Zürich), Dominik Brügger (catheter depth estimation, master thesis BFH, Biel), Simone Mortier (interpretation of isopotential maps, doctoral thesis Dr. med., University of Bern), Christof Baeriswyl (hardware implementation of algorithms, BFH, Biel), and Frédéric Waldmann (calculus for polynomials, master thesis, BFH Biel); they all clearly left their traces.

Further, there were many co-workers I worked with and each of them had contributed his or her valuable part to the project during the period of my thesis. So clearly did Thomas Niederhauser, who was my advisor in my beginning time at ARTORG, and who later set up and handled a certified quality management system at BFH, and managed the certified in-house catheter production according to current Swiss regulations (Swissmedic). Equally did Andreas Habegger (supervisor software design, supervisor ASIC die bonding process), Andreas Moser (catheter construction, software writing), Andreas Hugi (catheter design), David Metzger (software writing), Dominik Robellaz (non-uniform sample compression and interpolation, annotations), Emily Thompson (regulatory affairs, software writing, proofreading), Gerhard Kuert (catheter design), Luca Fleischmann (software writing), Lukas Bösch (ASIC design), Michael Nydegger (ASIC design), Nicolas Balmer (non-uniform sampling, data compression, and isopotential map analysing software), Rakesh Vasireddy (data analysis), Sandro Burn (electronics), Thanks Marisa (ASIC design), and Youyou Moser (electronic circuits).

The project also greatly gained from the valuable co-operation with the department of cardiology at the Bern University Hospital. In particular, my thanks go to Hildegard Tanner, Andreas Häberlin, and Romy Sweda. They performed the essential clinical trials to evaluate the performance of our tools and contributed valuable critical feedback.

As my new position at BFH overlapped with my thesis period by nearly one year, I had to handle additional teaching duties in parallel to the running project and the thesis writing. Fortunately, I could always rely on the support of Theo Kluter and Torsten Mähne whenever I

needed it, such as when I had to switch classes and the like, which finally saved me through this period.

Last but not least, I thank Stefan Moser for the weekly running at Zürich Fluntern, the discussions, and his honesty, farsighted advices, and his friendship.

And at very last but clearly most important, I thank my wife Yael for her—probably not fully voluntary, but unlimited—support over the past years, when she had to bear the main burden of our family commitments, next to her own demanding career.



# Abstract

Signals observed in biological systems originate most often from multiple, superimposed sources: some of those sources are producing continuously wandering baseline signals and are only vaguely known. By contrast, other signal sources are pulse-shaped and well characterized, for example, signals from neuronal or muscular cells, observed in electrocardiogram (ECG), electromyogram (EMG), and electroneurogram (ENG) signals, and the like. The challenges addressed in this thesis are to robustly identify and detect events in such biological signals, to separate sources from others, and to extract features. As biological signals take many shapes, the methods must be chosen carefully.

In this work, we contribute two robust signal processing methods, whereby both methods are based on localized, linear or non-linear, models: the first method uses localized autonomous linear state space models to identify and to detect events with certain characteristics; the second method applies localized cost functions of general polynomial forms to solve complex optimization problems. In this context, localization considers the signals only within a weighted window of finite or infinite length. Both methods, linear state space models and polynomial cost functions, are modest in their computational complexity and, hence, suitable for practical applications in wearable and, in particular, implantable medical devices.

This thesis is driven by a project in cardiology which conducts research in esophageal electrocardiography. As part of this project, we develop novel esophageal catheters with 3 dimensional electrode arrangements to localize cardiac events in 3-D space. The project has the overall goal to offer a superior, minimally invasive device for improved arrhythmia diagnostics.

In the first part of this thesis, we reconstruct the electrical field of the heart, as observed in the esophagus. Therefore, we apply our methods on measurements from an esophageal catheter solely. After depicting the results graphically as a 2 dimensional map, denoted as *esophageal*

*isopotential map*, we briefly discuss the medical implications of the obtained new modality. Furthermore, we derive a method to estimate the cardiac depolarization sequence on the heart surface (epicardium) of the left atrium. Solving such problems is also known as the *inverse problem of electrocardiography* and leads to the technique of *electrocardiographic imaging* (ECGI). The methods used to solve the problems in these first parts are derived in the following, second part of this thesis.

In the second part, we derive the methods required to solve the problems introduced in the first part. We introduce autonomous linear state space models (ALSSMs) and supplement them with local windows, which are generated by their own ALSSMs. We also combine multiple such ALSSMs, each localized by its own window, to generate more versatile models. We likewise superimpose multiple models of different time scales to discriminate signals of different temporal spread, and we join multiple models with adjacent windows to detect onsets of, or transitions between, multiple, alternating signal sources. Such joined models are either applied at particular time indices of interest, or repetitively over a whole signal leading to a sliding window filter. Such filters do not necessarily have scalar outputs, but rather provide signals of feature vectors. Applying appropriate transformations to such feature vectors simplifies unsupervised feature detection. Despite all these modifications and extensions on ALSSMs, we strictly preserve the inherent recursive computation rules of the involved linear state space models, and, thus, their efficient computations.

Further in the second part, we introduce cost functions of general polynomial forms and apply them to solve optimization problems. We expand their fields of application by introducing localization. To apply our method on a particular problem, we first project a given signal to a localized feature space, i.e., we locally approximate the signals by a given class of functions. Then, any further processing is executed in this lower-dimensional feature space. To efficiently handle problems of increased complexity, we also provide a new calculus. This calculus simplifies the manipulation of cost terms of polynomial forms.

In the third and last part, we apply our methods and provide solutions to the problems introduced. Finally, we conclude with a list of additional practical examples that we have already published, and which successfully apply our methods.

**Keywords:** Linear state space models; event detection and estimation; polynomial cost functions; unsupervised feature extraction.



# Kurzfassung

Signale wie sie in biologischen Systemen zu beobachten sind, sind meist das Ergebnis aus mehreren überlagerten Signalquellen mit jeweils unterschiedlichen Eigenschaften: Manche dieser Quellen produzieren sich stetig ändernde Grundlinien und deren Verhalten ist oft nur ansatzweise bekannt. Andere Quellen wiederum produzieren kurze, pulsformige Signale und sind gut charakterisiert; so zum Beispiel die elektrischen Aktivitäten von Neuronen und Muskelzellen. Solche Signale können unter anderem als Elektrokardiogramm (ECG), Elektromyogramm (EMG) oder als Elektroneurogramm (ENG) beobachtet werden.

Die Herausforderung in dieser Arbeit ist es, Ereignisse in solchen biologischen Signalen zuverlässig zu identifizieren, sowie verschiedene Signalquellen voneinander abzugrenzen und deren Merkmale zu extrahieren. Da biologische Signale in variierender Form in Erscheinung treten können, müssen die Analysemethoden entsprechend sorgfältig ausgewählt werden.

Wir haben in dieser Arbeit zwei robuste Signalverarbeitungsmethoden entwickelt. Beide Methoden basierend auf linearen und nichtlinearen, lokalisierten Modellen: Die erste Methode nutzt lokalisierte autonome lineare Zustandsraummodelle zur Identifizierung und Erkennung von Ereignissen mit bekannten Eigenschaften. Die zweite Methode nutzt lokalisierte Kostenfunktionen um komplexe Optimierungsprobleme zu lösen. Lokalisierung in diesem Kontext betrachtet Signale nur in einem gewichteten zeitlichen Fenster endlicher oder unendlicher Länge. Für solche Kostenfunktionen werden Polynome in allgemeiner Form verwendet. Beide Methoden sind moderat in ihrer Rechenkomplexität und damit prädestiniert für praktische Anwendung in mobilen und insbesondere implantierbaren Medizinalprodukten.

Diese Arbeit ist motiviert durch ein Projekt in der Kardiologie und betreibt Forschung in der Ösophagus-Elektrokardiographie. Im Rah-

men dieser Forschungsarbeit haben wir neuartige Ösophaguskatheter mit dreidimensionaler Elektrodenanordnung entwickelt, welche es erlauben, die Herzaktivität zeitlich und im 3-D-Raum zu erfassen. Diese räumliche Erfassung kardialer Ereignisse hat zum Ziel, ein minimalinvasives Medizinalgerät zur Verbesserung der Arrhythmie-Diagnostik zu entwickeln.

Im ersten Teil dieser Arbeit versuchen wir das elektrische Feld des Herzens einzig anhand von Messungen aus der Speiseröhre zu rekonstruieren. Anschliessend stellen wir die Rekonstruktion grafisch als zweidimensionale Karte dar, welche wir als *Esophageal Isopotential Map* bezeichnen. Weiter diskutieren wir die medizinischen Implikationen dieser neun Modalität. Zusätzlich leiten wir ein Verfahren her, welches die kardiale Depolarisationssequenz, die sogenannte *Activation Map*, auf Teilen der Herzoberfläche (Epikard) des linken Vorhofs rekonstruiert. Diese Rekonstruktion wird auch als *inverses Problem der Elektrokardiographie* bezeichnet und ist der Schlüssel zur *elektrokardiographischen Bildgebung* (*Electrocardiographic Imaging*, kurz: ECGI).

Im zweiten Teil werden die für die Aufgaben aus dem ersten Teil verwendeten Methoden hergeleitet. Hierzu führen wir autonome lineare Zustandsraummodelle (*Autonomous Linear State Space Models*, kurz: ALSSMs) ein und ergänzen sie jeweils mit einer lokalen Fensterung, welche selber als eigenes ALSSM beschrieben werden kann. Um die Flexibilität solcher Modelle zu erhöhen, kombinieren wir mehrere dieser ALSSMs verschiedener zeitlicher Ausdehnungen und durch verschiedene zeitliche Fensterungen lokalisiert. Damit lassen sich überlagerte Signale separieren. Werden verschiedene Modelle dieser Art in zeitlicher Folge aneinander gefügt, können damit auch Übergänge zwischen Signalquellen detektiert werden. Alle diese Modelle können entweder an einem spezifischen Zeitpunkt im Signal angewendet werden, oder aber wiederholt an jedem Zeitpunkt über die gesamte Länge eines Signals; dies entspricht einem sogenannten *sliding window filter*. Solche Filter haben dann nicht zwingend einen skalaren Ausgang, sondern können ganze Vektoren von Signalmerkmalen (*feature vectors*) generieren. Durch geeignete Transformationen solcher Vektoren eignen sie sich auch für eine unüberwachte Merkmalsextraktion (*unsupervised feature extraction*). Dies alles ist möglich unter Beibehaltung und Nutzung der effizienten rekursiven Berechnungsregeln, welche die zugrundeliegenden linearen Zustandsraummodelle ausmachen.

Ebenfalls im zweiten Teil werden Kostenfunktionen in der Form von allgemeinen Polynomen eingeführt. Solche Polynom-Kostenfunktionen werden verwendet um komplexe Optimierungsprobleme zu lösen. Da-

bei sind diese Kostenfunktionen wiederum lokalisiert, d.h. mit einer gewichteten zeitlichen Fensterung versehen. Um solche Kostenfunktionen auf ein konkretes Problem anwenden zu können, projizieren wir üblicherweise zunächst die gegebenen Signale auf einen lokalisierten Merkmalsraum (*feature space*), d.h. wir nähern die Signale durch eine gegebene Klasse von Funktionen, oft Polynome, an. Die anschließende Verarbeitung wird dann zugunsten der Effizienz nur noch in diesem niedrig-dimensionalen Merkmalsraum durchgeführt. Um auch Probleme von höherer Komplexität effizient bewältigen zu können, stellen wir eine Reihe eigener Rechenregeln zur Verfügung, welche die Umformungen von Polynom-Kostenfunktionen erleichtern. Solche Umformungen helfen auch den Echtzeitaufwand der Berechnungen zu minimieren.

Im dritten und letzten Teil wenden wir unsere entwickelten Methoden an und bieten damit Lösungen für die vorgestellten Probleme an. Schliesslich führen wir eine Liste bereits publizierter Beispiele auf, welche unsere Methode erfolgreich anwenden.

**Stichworte:** Lineare Zustandsraummodelle; Ereignisdetektion und Ereignisschätzung; Kostenfunktionen in Form von Polynomen; unüberwachtes Extrahieren von Merkmalen.



# Contents

|   |             |
|---|-------------|
| <b>Abstract</b>   | <b>vii</b>  |
| <b>Kurzfassung</b>  | <b>ix</b>   |
| <b>List of Symbols</b>  | <b>xvii</b> |
| <b>1 Introduction</b>   | <b>1</b>    |
| 1.1 Background . . . . .  | 1           |
| 1.2 Outline of Thesis . . . . .   | 2           |
| 1.3 Methods and Contributions . . . . .   | 4           |
| 1.4 Related Methods . . . . .   | 5           |
| <b>I 3-D Esophageal Electrocardiography</b>   | <b>9</b>    |
| <b>2 Physiology and Electrocardiographic Methods</b>                                      | <b>11</b>   |
| 2.1 Brief Summary of Cardiac Physiology . . . . .   | 11          |
| 2.2 Surface Electrocardiography (ECG) . . . . .   | 12          |
| 2.3 Esophageal Electrocardiography (EsoECG) . . . . .                                     | 13          |
| 2.4 Electrocardiographic Imaging (ECGI) . . . . .   | 15          |
| <b>3 The 3-D Esophageal ECG Project</b>   | <b>17</b>   |
| 3.1 The Esophageal Catheter EsoECG-3D . . . . .   | 17          |
| 3.2 Clinical Trials . . . . .   | 18          |
| 3.3 Results . . . . .   | 19          |
| 3.3.1 Esophageal Isopotential Maps . . . . .  | 19          |
| 3.3.2 Electrical Activity on the Left Atrial Posterior Wall<br>(Activation Map) . . . . . | 21          |

|           |  |           |
|-----------|--|-----------|
| <b>II</b> | <b>State Space Filters and Polynomial Costs</b>                            | <b>25</b> |
| <b>4</b>  | <b>Model-to-Signal Fit Using ALSSMs</b>                                    | <b>27</b> |
| 4.1       | Autonomous Linear State Space Models (ALSSMs) . . . . .                    | 28        |
| 4.2       | ALSSM Model Fitting . . . . .  | 31        |
| 4.2.1     | Squared Error Cost Functions . . . . .                                     | 31        |
| 4.2.2     | Model Fitting With Linear Constraints . . . . .                            | 32        |
| 4.2.3     | Recursive Computations and Window ALSSMs . . . . .                         | 33        |
| 4.2.4     | Weighted Samples . . . . .   | 37        |
| 4.2.5     | Time-Variant Windows . . . . .   | 38        |
| 4.2.6     | Multi-Segment Windows . . . . .  | 40        |
| 4.3       | The Hilbert Space Perspective . . . . .                                    | 41        |
| 4.3.1     | The Localized Hilbert Space $\mathcal{H}$ . . . . .                        | 41        |
| 4.3.2     | The Localized Feature Vector Space $\mathcal{F}$ . . . . .                 | 42        |
| 4.3.3     | An Euclidean Feature Vector Space . . . . .                                | 43        |
| 4.3.4     | The Feature Vectors as a Signal Surrogate . . . . .                        | 43        |
| 4.4       | The Feature-Vector Perspective . . . . .                                   | 44        |
| 4.5       | ALSSMs as LTI Filters . . . . .  | 44        |
| 4.6       | Localization of Events Using Local Cost Ratios (CRs) . . . . .             | 45        |
| 4.6.1     | Detection of Pulses . . . . .  | 47        |
| 4.6.2     | Detection of Model Switches . . . . .                                      | 47        |
| 4.6.3     | Detection of Outliers . . . . .  | 48        |
| <b>5</b>  | <b>Model-to-Model Fit Using ALSSMs</b>                                     | <b>51</b> |
| 5.1       | Fit of Any Two ALSSMs . . . . .  | 51        |
| 5.2       | Fit Using a Window ALSSM . . . . .   | 52        |
| 5.3       | Distance Between Two ALSSMs . . . . .                                      | 53        |
| 5.4       | Lossless Multi-Step Model Fitting . . . . .                                | 54        |
| 5.5       | ALSSM Signal Clustering in a Transformed Feature Space . . . . .           | 56        |
| <b>6</b>  | <b>Polynomial Cost Functions</b>   | <b>61</b> |
| 6.1       | Introduction . . . . .   | 61        |
| 6.2       | A Calculus for Polynomials . . . . .                                       | 63        |
| 6.2.1     | Vector Exponent Notation . . . . .   | 63        |
| 6.2.2     | Univariate Polynomials . . . . .   | 64        |
| 6.2.3     | Multivariate Polynomials . . . . .   | 67        |
| 6.2.4     | Table Summary . . . . .  | 72        |
| 6.3       | Applications . . . . .   | 75        |
| 6.3.1     | Squared Error Between Two Polynomials Within<br>a Given Interval . . . . . | 75        |

|                                      |   |            |
|--------------------------------------|---|------------|
| 6.3.2                                | Fit of Polynomials by Scaling in Amplitude and Time . . . . .             | 76         |
| 6.3.3                                | Offset-Free Fit of Polynomials by Scaling in Amplitude and Time . . . . . | 78         |
| 6.3.4                                | Delay Estimation Using Local Polynomial Approximations . . . . .          | 80         |
| <b>III Applications and Examples</b> |   | <b>85</b>  |
| <b>7</b>                             | <b>Activation Map Reconstruction From EsoECG Signals</b>                  | <b>87</b>  |
| 7.1                                  | Introduction . . . . .  | 87         |
| 7.2                                  | Solving the Inverse Problem With an ALSSM Filter Bank                     | 87         |
| 7.3                                  | Solving the Inverse Problem Using Polynomials . . . . .                   | 88         |
| 7.3.1                                | Method . . . . .  | 88         |
| 7.3.2                                | Physical Model . . . . .  | 90         |
| 7.3.3                                | Model Inversion . . . . .   | 92         |
| <b>8</b>                             | <b>Esophageal Isopotential Map Reconstruction</b>                         | <b>97</b>  |
| <b>9</b>                             | <b>Windowed State Space Filters</b>                                       | <b>99</b>  |
| <b>10</b>                            | <b>Further Applications and Examples</b>                                  | <b>105</b> |
| 10.1                                 | Signal Detection and Discrimination Using ALSSMs . . .                    | 106        |
| 10.2                                 | ALSSMs in a Filter Bank . . . . .   | 107        |
| 10.3                                 | Signal Interpolation and Reconstruction Using ALSSMs .                    | 107        |
| 10.4                                 | Applications of Polynomial Cost Functions . . . . .                       | 108        |
| <b>Appendices</b>                    |   | <b>111</b> |
| <b>A</b>                             | <b>Derivations in Chapter 4</b>   | <b>111</b> |
| A.1                                  | Derivation of (4.6) . . . . .   | 111        |
| A.2                                  | Derivation of (4.28) and (4.29) . . . . .                                 | 112        |
| A.3                                  | Derivation of (4.33), (4.33), and (4.34) . . . . .                        | 113        |
| A.4                                  | Derivation of (4.35), (4.37), and (4.38) . . . . .                        | 114        |
| <b>B</b>                             | <b>Derivations in Chapter 6</b>   | <b>117</b> |
| B.1                                  | Derivation of (6.7) . . . . .   | 117        |
| B.2                                  | Derivation of (6.43) . . . . .  | 117        |
| B.3                                  | Derivation of (6.65) . . . . .  | 118        |

---

|   |            |
|---|------------|
| B.4 Derivation of (6.37) . . . . .            | 118        |
| B.5 Derivation of (6.40) . . . . .            | 119        |
| B.6 Derivation of (6.58) . . . . .            | 119        |
| B.7 Derivation of (6.60) . . . . .            | 119        |
| B.8 Derivation of (6.71) . . . . .            | 120        |
| B.9 Derivation of (6.82) and (6.83) . . . . . | 121        |
| <b>Bibliography</b>                           | <b>123</b> |
| <b>About the Author</b>                       | <b>129</b> |



# List of Symbols

## Sets

|                  |   |
|------------------|---|
| $\mathbb{N}$     | $\{1, 2, 3, \dots\}$                      |
| $\mathbb{N}_0$   | $\{0, 1, 2, 3, \dots\}$                   |
| $\mathbb{R}$     | the set of real unumbers                  |
| $\mathbb{R}_+$   | the set of non-negative real numbers      |
| $\mathbb{R}_+^*$ | the set of strictly positive real numbers |

## Matrix and Vector Operations

|                        |  |
|------------------------|--|
| $I_n$                  | $n \times n$ identity matrix   |
| $\mathbf{1}_n$         | $n \times 1$ column vector of ones   |
| $\mathbf{0}_n$         | $n \times 1$ column vector of zeros  |
| $\ u\ $ or $\ u\ _2$   | L2 norm of a vector $u$  |
| $\langle u, v \rangle$ | standard inner product between two vectors $v$ and $u$                           |
| $\text{Tr}(A)$         | trace of a matrix $A$  |
| $\text{diag}(u)$       | diagonal matrix whose entries are the elements of vector $u$                     |
| $\text{vec}(A)$        | column vector obtained by stacking columns of a matrix $A$ on top of one another |
| $\text{vdiag}(u)$      | a short form of $\text{vec}(\text{diag}(u))$                                     |
| $\otimes$              | Kronecker product  |
| $\circ$                | Hadamard product (element wise product)  |
| $a_{\{n\}}$            | $n$ th element in vector $a$   |
| $A_{\{n,m\}}$          | element in $n$ th row and $m$ th column of matrix $A$                            |

## Abbreviations and Acronyms

|        |                                     |
|--------|-------------------------------------|
| ALSSM  | Autonomous Linear State Space Model |
| ECG    | Electrocardiogram                   |
| EsoECG | Esophageal Electrocardiogram        |
| LSSM   | Linear State Space Model            |

# Chapter 1

## Introduction

### 1.1 Background

This thesis is driven by the *3-D Esophageal ECG* project, an interdisciplinary project led by the *Institute for Human Centered Engineering, Bern University of Applied Sciences*, Switzerland. This project conducts research in esophageal electrocardiography and aims to develop a superior esophageal electrocardiographic device, improving the arrhythmia diagnostics in cardiology. The key component of this device is a novel esophageal catheter denoted as *EsoECG-3D*, which is inserted through the patient's nose into the esophagus to its target position behind the patient's heart. The signals measured with this catheter offer a spatial view of the cardiac electrical activity thanks to the catheter's 3 dimensional electrode arrangement and the beneficial anatomical location of the catheter next to the heart's atria.

The catheter *EsoECG-3D* was designed and fabricated at the Bern University of Applied Sciences under an ISO 13485<sup>1</sup> certified environment. Clinical trials involving patients and healthy subjects were subsequently performed in collaboration with the *Department of Cardiology, Bern University Hospital*, Switzerland, and are fully approved by the local ethical authorities and the *Swiss Agency for Therapeutic Products, swissmedic*. The signal processing research was conducted at the *Signal and Information Processing Laboratory (ISI), ETH Zürich* under supervision of Prof. Hans-Andrea Loeliger.

---

<sup>1</sup>ISO 13485 is a standard, representing the requirements for a comprehensive quality management system for the design and manufacture of medical devices

## 1.2 Outline of Thesis

This thesis is organized in three parts: Part I identifies particular diagnostic needs in rhythmology, a special field in cardiology. Part II develops new signal processing ideas and methods, which are finally applied in Part III to provide solutions to the needs identified in Part I.

### Part I: 3-D Esophageal Electrocardiography

Part I starts in Chapter 2 with an overview of electrocardiographic diagnostic methods used in cardiology, along with their current limitations. Such limitations are (among others): the lack of resolution in atrial signals and the need for non-invasive methods to study the cardiac depolarization sequence of the heart muscle, also known as the *electrocardiographic imaging* (ECGI) methods. In Chapter 3, we introduce our project performing 3-D electrocardiography and already give an overview of the new results achieved from a medical perspective.

### Part II: State Space Filters and Polynomial Costs

Part II contributes new signal processing solutions and extends existing concepts in order to address the problems introduced in Part I.

For signal detection, discrimination, and filtering tasks, we introduce in Chapter 4 autonomous linear state space models (ALSSMs) and supplement them with local windows, which are generated by their own ALSSM. We also combine multiple ALSSMs, each localized by its own window, to obtain more versatile models and windows. We likewise superimpose multiple models of different time scales to discriminate pulses from their baseline signals, and we join multiple models with adjacent windows to detect either the onset of a model or the transitions from one model to another. Such joined models are either applied at particular time indices of interest, or repetitively over a whole signal, leading to a sliding window filter. Such filters do not necessarily have scalar outputs, but rather provide signals of feature vectors. Finally, despite all these modifications on ALSSMs, we strictly preserve the inherent recursive computation rules of the involved linear state space models, and, thus, their efficient computations.

For signal clustering and unsupervised feature detection, we introduce in Chapter 5 a distance measure for ALSSMs in an Euclidean feature space. In this space, we apply common clustering methods and thereby blindly extract features of single- and multi-channel signals.

To efficiently solve optimization tasks, we introduce in Chapter 6 cost functions of general polynomial forms and expand their field of application by introducing localization. In this context, localization refers to a local perspective, where the signal is only considered within a window of finite or infinite length. For that, we initially project any observed signal to a localized feature space, i.e., we locally approximate the signals by polynomials using a window. Then, any further processing is exclusively done in this feature space. To efficiently handle such optimization problems of increased complexity, we also provide a new calculus for polynomials. This calculus simplifies the manipulation of cost terms of polynomial forms and, e.g., supports the separation of observation-independent terms to minimize the real-time computation efforts.

The presented methods in Part II are motivated by the practical needs from Part I. Nevertheless, we think that our approaches are far more general to be used in just our target applications: the derived methods are modest in their need of computation power and memory usage, as is preferred for wearable and implantable medical devices, and likewise for mobile applications in general.

## Part III: Applications and Examples

Part III applies the methods derived in Part II on the medical applications introduced in Part I and other practical problems. In a first application in Chapter 7, we use cost functions of polynomial forms to solve the inverse problem of electrocardiography for our esophageal ECG measurements. In a second application in Chapter 8, we refer to our published method extensively applying polynomial cost functions to track the position of an esophageal catheter inside the body and to reconstruct so-called esophageal isopotential maps of high atrial resolution. We here supplement the published method with details on the computations applying our calculus. In a third application in Chapter 9, we summarize our framework for signal detection and discrimination based on localized autonomous linear state space models. This framework is already published in [47]. Finally, in Chapter 10, we conclude Part III with an overview of applications and examples based on our methods that are already published. These examples include P wave and QRS complex detection in ECG signals, baseline reconstructions, detection of modulated pulses, signal smoothing, and others.

## 1.3 Methods and Contributions

In this thesis, we search for feasible methods that meet the current state of the art for analyzing biological signals such as electroneurograms (ENGs), electromyograms (EMGs), electroencephalograms (EEGs), and electrocardiograms (ECGs) including the signals from our 3-D Esophageal ECG project. We look for robust methods that perform in practical working environments and meet the constraints of medical devices, such as reliability, limited energy supply, space restrictions, and more. Although it is not a primary goal for the time being, we are also interested in methods that are suitable for hardware algorithm implementations, e.g., in an FPGA (Field Programmable Gate Array) or even in an ASIC (Application Specific Integrated Circuit) to further minimize the power consumption, as low power consuming hardware is crucial for portable and implantable medical devices.

In this context, we explore the following methods: linear state space models in a least squares framework and general cost functions of polynomial forms to state and solve optimization problems.

**Autonomous Linear State Space Models** By contrast to most statistical approaches, we only use *autonomous* linear state space models (ALSSMs), which do not have an input and whose outputs are fully determined once an initial state is set. The use of ALSSMs leads to efficient recursions for the computation of squared error cost functions. ALSSMs are convincing for our applications since:

- an ALSSM supplemented by an appropriate window can localize events like those often observed in biological signals,
- the ALSSM system order required to detect and discriminate events is usually low (in comparison to the number of discrete-time samples involved) and independent of the actual sampling rate,
- ALSSMs commonly have a continuous-time equivalent and therefore facilitate the modeling of non-uniformly sampled or continuous-time signals,
- the computation of ALSSMs is non-iterative, i.e., in most applications a single forward or/and a single backward recursion is sufficient,
- and finally, the ALSSM methods as proposed in this thesis, scale linearly with the number of samples and are of low complexity.

**Polynomial Cost Functions** We observed that localized polynomials, i.e., polynomials over a certain interval and supplemented with a window, are convenient to locally approximate many type of pulse-shaped signals. This is not a new observation, since there are already signal processing methods proposing this, e.g., when using splines [44].

To go further, we use our approximations of polynomial forms to define problem-specific cost functions: since most cost functions use the measure of squared errors—which is by definition a second order polynomial form—cost functions based on polynomial approximations can also be transformed into terms of convenient polynomial forms.

We explored this idea on multiple problems of varying complexity and observed that, despite similarities, each problem statement requires its own algebraic transformation to gain a “convenient” polynomial form. These transformations are performed manually and for any non-trivial example get exhaustive despite using only common linear algebra. Therefore, we derived our own calculus for polynomials to ease this manual work. Overall, we found localized polynomial cost functions to be convincing for our applications since:

- coefficient terms are fully or in parts precomputable offline,
- remainders to be processed and minimized in real time are polynomial or rational functions,
- minimization of polynomials is done efficiently in closed form or with well-known numerical methods,
- and, thanks to its low remaining real time complexity, the method is suitable for low-power and low-memory hardware applications.

## 1.4 Related Methods

There is a wide variety of existing methods to process biological signals. Most techniques use some combination of linear filtering, including, but not restricted to, Fourier transforms [45], matched filters [43], wavelets [33], or filter banks [1]. However, all these methods have limitations, especially when interferences and irregular baselines cannot be neglected.

Another suitable approach is the use of linear dynamical systems for signal modeling with occasional model switches [14, 16, 31]. Yet, such models require many parameters to be estimated and their estimations are complex. A related approach was recently proposed in [30, 54], where sparse jumps in linear state space models are handled with normal priors

of unknown variances. These algorithms are iterative and computationally demanding.

There are many practical examples from the field of medical research. An overview of signal processing approaches used for QRS wave detection in ECG signals is given in [25], a selection of ECG baseline removal techniques in [27], or an overview of spike detection in electroencephalogram (EEG) signals in [51]. Some of the methods used suffer from undesired signal distortion or insufficient pulse shape or onset point preservation, which might have a direct impact on the diagnosis, such as when measuring the ST segment elevation on suspicion of a myocardial infarction.

Overall, biological signals show a wide variability in signal characteristics and are most often a superposition of various components. In many processing tasks, single such components need to be detected and separated from other components, with only minimal distortion added to the signal; in particular the preservation of pulse-shapes and onset-points is essential.

Our first approach presented in this thesis mainly borrows ideas from autonomous linear state space models and from recursive least square computations with localization, i.e., with using appropriate windows. Our second approach solves optimization problems by minimizing localized cost functions of general polynomial shapes.

**(Autonomous) Linear State Space Models** Linear state space models have long been used in different fields such as control systems [22–24] and more recently also in signal analysis and machine learning [5, 15, 52]. In contrast to general linear state space models, input-free, i.e., autonomous, linear state space models are rarely used in practical applications so far; but they have been thoroughly studied at our institute in the past [8, 39, 53, 54, 56]. The autonomous linear state space models were thereby (mainly) used in a squared error setting, borrowing the idea of recursive least squares to gain efficient computation rules. Recently, linear state space models were also used with higher-order cost functions, i.e., beyond squared errors [55].

**Recursive Least Squares** The concept of recursive least squares is seen in a wide variety of applications including adaptive filters, Bayesian filters, and Kalman filters [18, 24, 26]. Applying recursive least squares on autonomous linear state space models was studied intensively at our institute over the past years [8, 39, 53, 54, 56].



**Localization Using Windows** Localization using windows is an old idea. Often seen are windows of one-sided or two-sided exponentially decaying shapes as, for example, proposed in [18,26]; in our institute, they are used in multiple applications [53,54,56]. Rectangular windows involving only a fixed number of samples are also common [6]; in combination with the concept of recursive least squares, these windows lead to filters referred to as *fixed-memory* or *limited-memory filters* [4, 9, 20, 40, 41]. Finally, Bruckstein et al. proposed in 1985 limited memory filtering for signals with given state space representation [7].

**Likelihood ratios** *Likelihood ratios* or also *log-likelihood ratios* are well known from statistical theory, and are interpreted as the ratio of the probabilities between two hypotheses. At our institute, localized versions of such model likelihoods were used to detect certain events in given observations [29], [56, Section 7.3], [39, Part II].

**Polynomial Cost Functions** Polynomial cost functions of second order are ubiquitous in signal processing, and more general, in all engineering and natural sciences. Least square methods or statistical signal processing using Gaussian distributions are probably the most popular examples. All these methods rely on quadratic costs (or their Gaussian counterparts), but quadratic cost terms are not always suitable, which has motivated the extension to non-linear approach such as extended Kalman filter (EKF) or unscented Kalman filter (UKF) [21], particle filters, and others [11]. The use of cost functions of general polynomial forms seems an obvious generalization to quadratic cost functions, at the cost of losing the immediate statistical counterpart, the Gaussian distribution, and often also at the cost of losing its closed form solutions.

**Signal Approximation and Interpolation by Polynomials** Already in 1946, Schoenberg proposed to approximate signals by piecewise polynomials, i.e., splines [42]. Schoenberg also introduced the now well-known B-splines, a beneficial representation of splines, which are computationally easy to handle [12]. In the fields of mechanics and computer aided design, splines are extensively used for surface modeling [13]. Also in signal processing, there is a wide field of applications for polynomials [44]. So, splines are often applied to interpolate equidistant or non-equidistant sampled signals.



## Part I

# 3-D Esophageal Electrocardiography



# Chapter 2

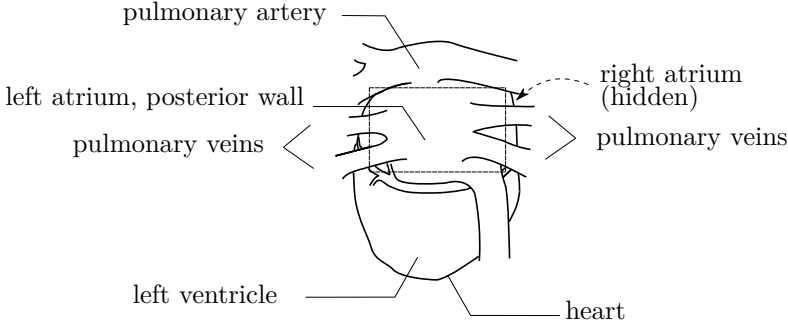
# Physiology and Electrocardiographic Methods

In this chapter, we provide a brief summary of cardiophysiology, followed by the summary of some of today's electrocardiographic methods.

## 2.1 Brief Summary of Cardiac Physiology

The cardiac activation denotes the spreading of the electrical cell depolarization process through the heart muscle with respect to time. A charged myocardial cell has an intracellular potential of approximately  $-70$  mV, a discharged, i.e. a depolarized cell, a potential of about  $0$  mV. The switch of a single cell is hardly recordable on the body surface or in the esophagus, but the orchestrated overturning of many cells at the same time can indeed be observed; this yields the well-known electrocardiogram signal of about  $1$  mV amplitude on the body surface as well as in the esophagus.

In a healthy human heart the atrium is triggered by the sinus node, which is located in the right apical atrium (see also the anatomical view of the human heart in Fig. 2.1). Subsequently, the ventricle gets activated via the atrioventricular node (av node) with a delay of approximately  $100$  ms. This av node interconnects the atrium and the (otherwise electrically isolated) ventricle. The current front line of this depolarization process is denoted as the *activation front* (cf. Fig. 2.2).



**Figure 2.1:** Back view (dorsal view) of a human heart, including the outlets of the main vessels. Our EsoECG-3D catheter analyzes the electrical activity on the posterior wall of the left atrium (dashed box). The right atrium is mostly hidden behind the left atrium.

## 2.2 Surface Electrocardiography (ECG)

The 12-lead electrocardiogram (ECG) has been the most important diagnostic tool in cardiology for more than half a century. This standard method of ECG recording uses 10 electrodes<sup>1</sup> placed on predefined anatomical spots on the patient's thorax and limbs, and is the method of choice to identify cardiac problems. Nevertheless, for many questions in rhythmology<sup>2</sup> this method shows insufficient resolution, in particular in the heart's atrial regions, despite the atria causing many of the common cardiac arrhythmias such as atrial flutter, atrial fibrillation, and many others. The atrial signal in a surface ECG, denoted as the P waves, is about 5 to 10 times smaller in amplitude than the ventricular signal, the QRS complex. This difference in amplitude is caused by the discrepancies in myocardial muscle mass: while the atrial myocardium measures only a few millimeters, the thickness of the ventricle muscle is more than 10 mm. In addition, the atria are unfavorably located for surface ECGs; they are deep in the thorax, far from the surface electrodes. Despite these limitations, the importance of ECGs in cardiology is undoubtable.

<sup>1</sup>A so-called 12-lead ECG only uses 10 physical connections and with these, only 8 independent signals are measured. Nevertheless, a clinical 12-lead ECG displays 12 ECG signals, whereof 4 signals are redundant.

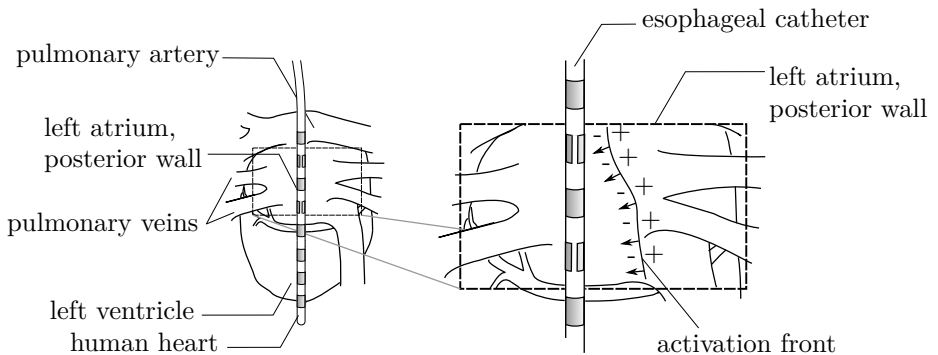
<sup>2</sup>Rhythmology is the medical field studying and curing the electrical cardiac conduction system

## 2.3 Esophageal Electrocardiography

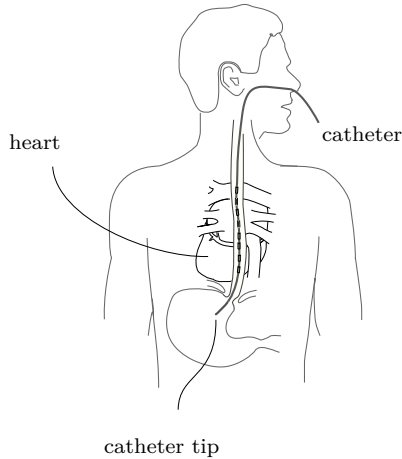
An alternative low-risk electrocardiogram recording method showing a high atrial resolution is the esophageal electrocardiography (EsoECG). EsoECG is an old and well-known technique, recording the cardiac electrical signal inside the esophagus, i.e., it measures the electrical potentials directly on the mucosa surface.

First experiments with esophageal electrocardiography started around 1906 [10], but until today, it remained a niche technology. There might be various reasons for this: probably the surface ECG was sufficient for the questions at that time and with increasing experience in this method, it became difficult for any alternative method to compete.

Today, thin and flexible catheters are used, equipped with multiple electrodes at the far end of the catheter. These catheters are inserted through the nose or the mouth into the patient's esophagus, until the catheter is located directly behind the heart (see Fig. 2.2 and Fig. 2.3). At this location, the left atrium is only a few millimeters and the right atrium some centimeters from the esophagus [3, 36, 50]. The esophageal



**Figure 2.2:** The left figure illustrates a dorsal view (i.e. a view from behind) of a human heart with an esophageal catheter located inside the esophagus (the esophagus is not shown). The right figure illustrates a detail view of the left atrial posterior wall and the catheter. Further, an example of an activation front, the myocardial depolarization process, is depicted: the arrows are pointing in the propagation direction of the activation front with the depolarized + cells on one side and the polarized – cells on the other side.



**Figure 2.3:** View from the back side (posterior view) of a human heart in situ. The esophageal catheter is placed through the nose into the esophagus, such that the electrodes are located close to the back (dorsal) side of the heart.

catheter there observes the electrical activity of the atria with high signal quality. (cf. Fig. 2.2)

Nevertheless, esophageal electrocardiography has its own limitations: first, esophageal catheters are not fixed to the surrounding mucosa, but are steadily floating back and forth due to the patient's breathing and body motions. Thus, the recording location is continuously altering. Second, EsoECG signals show strong interferences, mainly caused by the continual changes of the electrochemical environment inside the esophagus, such as changes in the saliva composition, including acid reflux from the stomach. Third, the space for any device inside the esophagus is limited to its tube-like interior and restricts the possible electrode arrangements, i.e., it is not feasible to widely spread the electrodes apart as we do on the thorax surface for standard ECGs.

Nowadays, materials and techniques have greatly improved and, if needed, local surface anesthesia is used to avoid discomfort during the insertion procedure. Furthermore, the advances in signal processing, together with the rapidly increasing computing power, have opened up new possibilities. Therefore, it makes good sense to look again at esophageal electrocardiography from today's perspective.



## 2.4 Electrocardiographic Imaging (ECGI)

The idea of *electrocardiographic imaging* (ECGI) with the overall goal to characterize the electrical activity of the heart, has been around for several decades. ECGI mostly uses surface ECG recordings alone, but there are some cases where these signals are supplemented with some intra-cardiac (from inside the heart) or esophageal measurements. However, of main interest are the non-invasive ECGIs, i.e., mapping methods where no additional intra-cardiac measurements are required. There was recently a strong upsurge to use this method, most likely driven by the increasing computation power available and by new therapeutic options requiring precise diagnostics [19]. ECGI usually uses an individual anatomical volume conductor model derived from an individual computed tomography scan (CT scan), and a mathematical model for the electrical propagation through the myocardium.

Overall, ECGI has made enormous progress over the last years, but either relies on personalized anatomical models (CT scans) or is strongly limited in its atrial resolution.



# Chapter 3

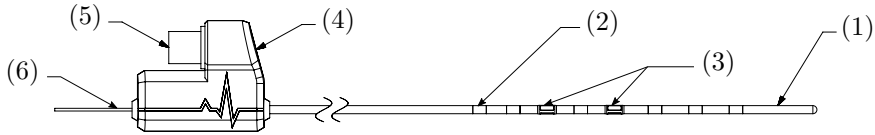
## The 3-D Esophageal ECG Project

This chapter provides an insight into the 3-D Esophageal ECG project and presents the first results from a medical perspective.

### 3.1 The Esophageal Catheter EsoECG-3D

To lift esophageal electrocardiography to today's technology standards, a catheter with a 3-dimensional electrode configuration [46] was developed at Bern University of Applied Sciences in 2016. This catheter, named EsoECG-3D, was designed in parallel to this work and is approved by Swissmedic authority for the use in clinical trials. EsoECG-3D is based on a soft 3 mm diameter polyurethane tube, where the caudal end, the end which is inserted into the body, is sealed and equipped with a total of 14 electrodes. Of these 14 electrodes, 8 are common ring electrodes and 6 electrodes are small ring segments, arranged in 2 groups of 3 electrodes in each group, and each group spanning around the catheter with a 120 degree offset per electrode. The distance between the electrodes along the catheter varies between 10 mm in the upper (cranial) part, and 15 mm in the lower (caudal, towards the catheter tip) part. The cranial end holds a connector as shown in Fig. 3.1.

This EsoECG-3D catheter was applied for the first time during a collaborative clinical trial in 2018.



**Figure 3.1:** Esophageal catheter, EsoECG-3D, designed and fabricated at Bern University of Applied Sciences. (1) catheter tip of 3 mm soft polyurethane tube; (2) ring electrode; (3) split 3-D electrodes; (4) connector housing; (5) connector; (6) guide wire (catheter insertion aid).

## 3.2 Clinical Trials

This thesis uses data sets from two clinical trials with 20 volunteers in the first, and 52 in the second. Both trials were performed in collaboration with Bern University Hospital, Switzerland.

**Clinical Trial 2016, MC-EECG (n=20)** We performed a first clinical trial in 2016, using commercially available esophageal catheters (*Esoflex 10S*, *FIAB SpA*, *Italy*) with 10 ring electrodes with 1 cm spacing. With these catheters, we have recorded EsoECG signals from 14 patients with atrial arrhythmias, such as atrial fibrillation and atrial flutter, and from 6 healthy subjects. This trial was approved by the Swiss ethics commission (No. 149/15; [ClinicalTrials.gov/NCT02541175](https://clinicaltrials.gov/NCT02541175)).

**Clinical Trial 2018, esoECG-3D (n=52)** Our second clinical trial was a *first in man* trial for our own catheter, EsoECG-3D, involving a total of 52 volunteers, whereof 12 were healthy volunteers and 40 were patients with a history of cardiac diseases. The patients underwent a scheduled cardiac intervention (electrophysiological surgery) due to preexisting heart disorders such as atrial fibrillation, atrial flutter, and others. For these patients, our catheter recorded the esophageal signals in synchronization to the intra-cardiac measurements obtained during the interventions; these intra-cardiac measurements are the accepted “ground truth” signal in that field and are taken as the reference to validate our results. The 12 remaining volunteers had no history of cardiac disorders and, therefore, the esophageal measurements were recorded independent of any intervention. Our catheter as well as the clinical trial were approved by Swissmedic authority and Swiss ethics commission ([ClinicalTrials.gov/NCT03365440](https://clinicaltrials.gov/NCT03365440)).

## 3.3 Results

### 3.3.1 Esophageal Isopotential Maps

Esophageal isopotential maps, EsoIPMs, are a new method of cardiac field representation and depict the dynamical cardiac electrical field of the heart as observed in the esophagus. EsoIPMs are reconstructed using our own processing methods, applied on our clinical data sets. We already published a basic version of this reconstruction technique in [50] in 2018, along with results based on our first clinical trial performed in 2016. The medical implications of this methods were studied by Simone Mortier [35] and we presented the results also at the *Venice Arrhythmias Conference 2018* [49].

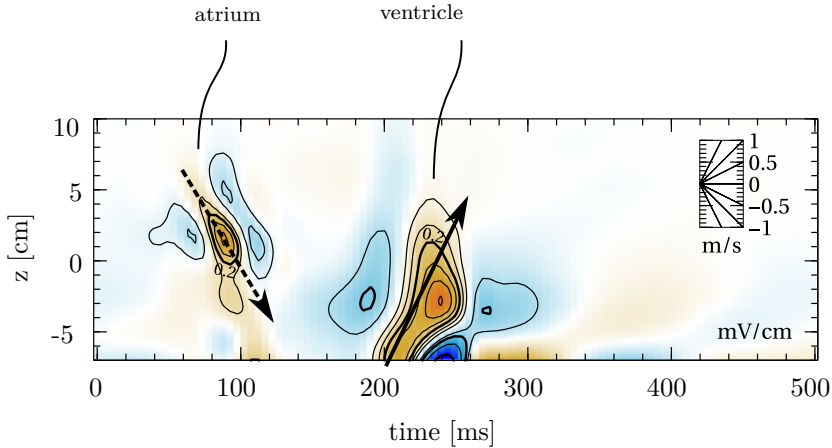
In the following, we give a brief summary of this published method. Next, we extend the method in a more efficient way to solve the problem discussed in [50] and to perform the corresponding computations, before we conclude with some additional findings.

#### Summary

EsoIPMs plot the temporal changes of the electrical field recorded in the esophagus; the electrical field is thereby reduced to the single dimension along the esophagus' longitudinal axis. Such an EsoIPM is then plot in a 2 dimensional manner with the time on the horizontal axis, the location along the esophagus on the vertical axis, and the measured electrical potential displayed as electric isopotential lines. A first example of an EsoIPM is given in Fig. 3.2.

The reconstruction of such EsoIPMs from the recorded esophageal ECG signals mainly has to deal with two difficulties: first, the number of available measurement electrodes (channels) is rather low and leads to undersampling of the electrical field in space. Second, the constant drift of the catheter device inside the esophagus, as mentioned in Section 2.3, constantly alters the location of the electrodes within the cardiac field of interest. Figure 3.3 depicts such an electric field and illustrates the moving sampling locations along the esophagus over time.

Interestingly, finding a solution for the latter, the moving electrode positions problem, also helps to solve the former, the undersampling problem. The catheter's arbitrarily drifting back and forth along the esophagus axis results in a spatial sampling with small but "randomly" varying changes of the electrode positions. As we have shown in [50], we



**Figure 3.2:** Example of an EsoIPM, displaying the reconstructed cardiac electric field, observed along the esophagus and displayed as a two-dimensional (2-D) contour plot. An EsoIPM shows the position along the esophagus on the vertical axis  $z$  and the time evolution on the horizontal axis. The 2-dimensional plot is additionally encoded in colors according to the electric field's strength (positive values in orange/red, negative values in blue). The atrial signal in this example shows a propagation from top to bottom (i.e., cranio-caudal, dashed arrow), while the propagation in the ventricle runs from bottom to top (i.e., caudo-cranial, solid arrow). Figure source: [50].

can take advantage of this random catheter drift to increase the spatial signal resolution.

## Method

The reconstruction of EsoIPMs applies our method of polynomial cost functions, which is introduced later in this thesis in Chapter 6.

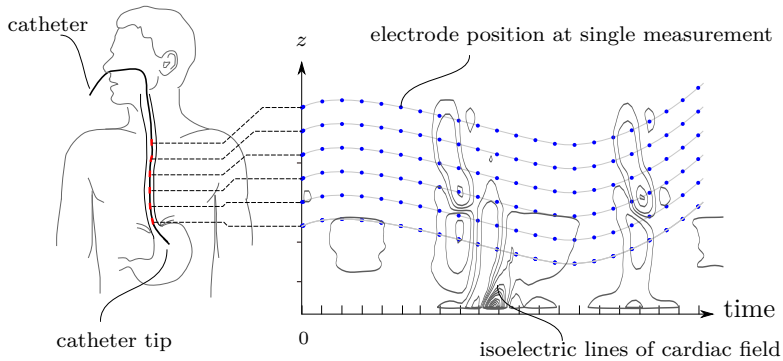
## Examples

Figure 3.2 shows a first example of an EsoIPM, reconstructed using EsoECG signals from a healthy subject. This EsoIPM clearly discriminates and localizes graphically the atrial and ventricular activities. Figure 3.4 is a pathologic example from a 67 year old patient with atrial

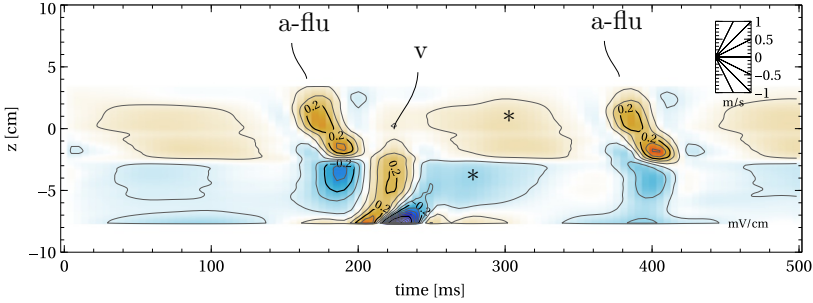
flutter, i.e., where the atrial electrical activity circles endlessly around the heart's atria. Such a diagnosis would clearly indicate for a therapy, most likely for an intra-cardiac catheter ablation therapy, with which the aberrant conducting path gets isolated. Figure 3.5 shows the same sequence as Fig. 3.4 but over a longer time period.

### 3.3.2 Electrical Activity on the Left Atrial Posterior Wall (Activation Map)

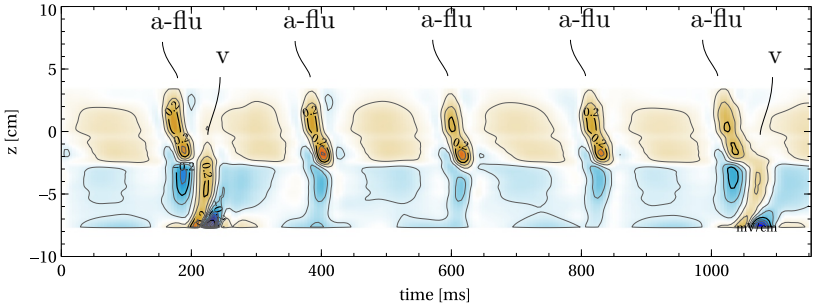
A cardiac activation map is a well established tool in cardiology and graphically depicts each single cardiac depolarization process of a pa-



**Figure 3.3:** Illustration of an inserted esophageal catheter and the observed dynamic cardiac electric field overlain by the wandering electrode positions: the illustration on the left shows a catheter with 6 electrodes (red marks) inserted into the esophagus. The plot on the right shows a 2-dimensional plot with the position along the esophagus on the vertical axis, denoted as  $z$ -axis, and the time on the horizontal axis. In this plot are the effective position of each single electrode over a short sampling time period marked (blue dots). We note that the actual insertion depth ( $z$ -axis) of the single electrodes continuously alters due to patient's movements and breathing. Further overlaid onto this plot are the lines of isoelectric fields (thin black curves) inside the esophagus.



**Figure 3.4:** EsoIPM of a 67 year old patient with typical counterclockwise atrial flutter: atrial flutter waves (a-flu), ventricular wave (v), baseline signal artifacts (\*). Figure source: [50].



**Figure 3.5:** EsoIPM as in Fig. 3.4 but with time axis spanning 1200 ms to emphasize the heart rhythm (counterclockwise atrial flutter with 4:1 conduction from the atrium to the ventricle): atrial flutter waves (a-flu), ventricular waves (v). Figure source: [50].

tient. Data sources for the reconstruction of such maps are commonly invasively measured electrical potentials, gained using intra-cardiac catheters navigated into the heart through a vein or an artery.

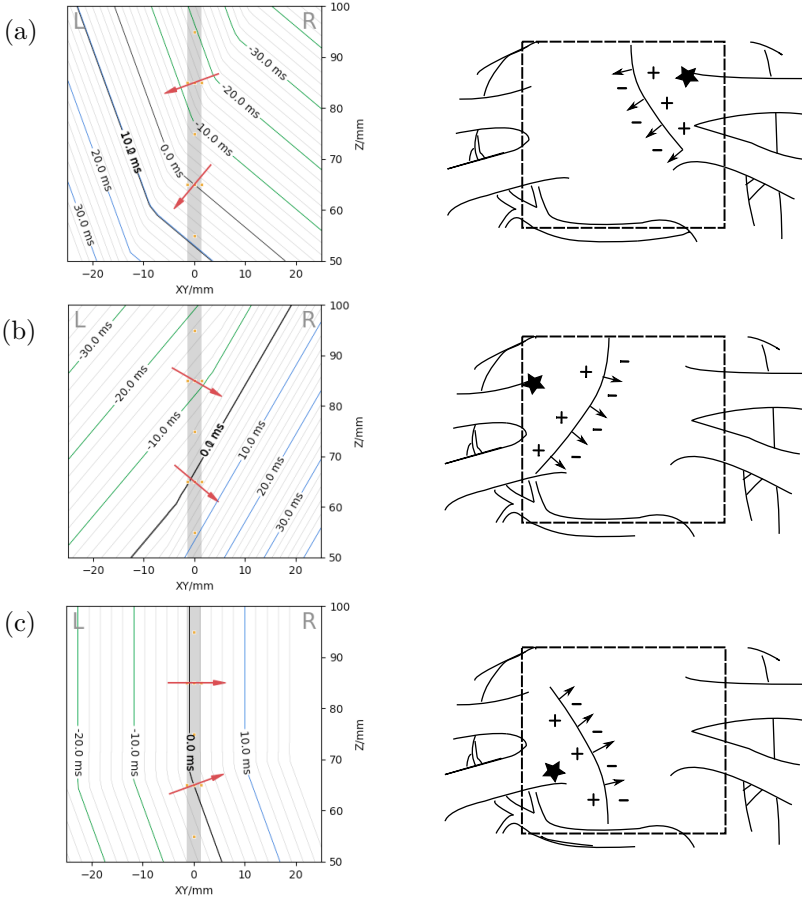
Nowadays, non-invasive ECGs are increasingly becoming a viable source for such activation maps, denoted as Electrocardiographic Imaging (ECGI), cf., Section 2.4. A high-quality map usually immediately leads to a diagnosis and thus defines the therapy options. Therefore, a reliable and preferably non-invasive method to reconstruct these activation maps is of high interest in rhythmology.



While common ECGs are based on measurements from dozens to hundreds of electrodes spread over the thorax, the EsoECG-3D catheter only offers 14 electrodes measuring in the esophagus. To make matters even worse, the spatial spreading of the electrodes in our catheter in the plane perpendicular to the catheter is below 3mm. Nevertheless, we found that ECGI is possible based on EsoECG-3D records alone, if we restrict to heart areas close to the catheter, in particular to the left atrium. Furthermore, as our catheters measure atrial signals with a high signal-to-noise ratio, we can do this inversion for this particular area based solely on general models, solving the *inverse problem* for the left atrial posterior wall, cf. Fig. 2.2. Note that the signal processing aspects of our method are discussed in details later in Chapter 7.

### **EsoECG-3D Activation Maps: Preliminary Results**

These results are based on the data sets recorded during the second clinical trial, in 2018. In this trial, patients were electrically stimulated (paced) at well-defined locations inside the heart (similar to stimulations by a cardiac pace maker). Figure 3.6 demonstrates 3 different such pacing locations for a single patient and their resulting activation maps, which were reconstructed using solely esophageal measurements. We clearly note, that the propagation direction on the posterior wall of the left atrium fully agrees with the expectation according to the pacing location: if we pace at the left upper pulmonary vein, then the propagation spreads centrifugal from that spot and leaves its mark in our activation map on the left atrial wall. For other trigger locations, like at the right upper or right lower pulmonary veins, at the atrial roof, or at other anatomical structures, the propagation spreads accordingly. We emphasize at this point that for the presented results no individual anatomical data sets such as CT or MRI scans are used.



**Figure 3.6:** Reconstructed activation maps for patient IDA07 from our clinical trial. The reconstruction shown on the left-hand side uses our method based solely on esophageal measurements. The patient shown was artificially paced using an intra-cardiac pacemaker electrode placed at (a) the right superior pulmonary vein (RSPV), (b) the left superior pulmonary vein (LSPV), (c) the left inferior pulmonary vein (LIPV). These pacing locations are marked as  $\star$ . The estimated activation maps (shown on the left-hand side) use isochronous lines, indicating the location of the activation front at  $t = -30$  ms,  $-20$  ms,  $\dots$ ,  $30$  ms. Time  $t$  is a relative measure and  $t = 0$  ms indicates when the activation front passes closest to the catheter.

## Part II

# Windowed State Space Filters and Polynomial Cost Functions



## Chapter 4

# Model-to-Signal Fit Using ALSSMs

Parametrized models are commonly fit to signals or data sets by minimizing a squared error cost function. While this is computationally feasible for small data sets, it rapidly becomes computationally expensive for larger signals and in particular when a moving time domain is involved, e.g., when the fit needs to be repetitively performed at many locations along the signal, as when applying a sliding window. Moving time domains or sliding windows typically require a recomputation of the fit at each location. Here, the use of autonomous linear state space models (ALSSMs) drastically reduces the computing effort due to their recursive definition; the inherent recursions of ALSSMs directly lead to recursive forms of squared error computations, which is indeed a form of recursive least squares.

Recursive least squares (RLS) error computation was already used by Gauss at the beginning of the 19th century. In present day, the concept of RLS is seen in a wide variety of applications including adaptive filters, Bayesian filters, and Kalman filters [18,24,26]. Applying RLS on ALSSMs is not yet very common, but has been studied intensively at our institute over the past years [8,39,53,54,56].

Filters involving only a fixed number of samples, later on also referred to as *fixed-memory* or *limited-memory filters*, were already proposed by Blum in 1957, where RLS was used to efficiently fit polynomials to a finite set of data within a moving window [6]. Savitzky and Golay have proposed in 1964 to use general convolutional filters or least squares fits

over a fixed number of samples to remove random noise from discrete-time data sets [40, 41]. Jazwinski introduced in 1968 a filter related to maximum likelihood estimates with limited memory length in a sliding window manner, called a *limited memory filter* [20]. Buxbaum and Bierman et al. continued on recursive fixed-memory least-squares filtering in 1975 and 1976, and published a fixed-memory Kalman filter version [4, 9]. Finally, Bruckstein et al. proposed in 1985 limited memory filtering for signals with given state space representation [7].

In this chapter, we explore ALSSMs in a least-squares setting and introduce sliding window filters based on ALSSMs, which either consider a fixed number of observations per estimate, i.e., *limited-memory filters*, or an unrestricted number of observations with fading weights. In both cases, we provide efficient recursion rules, in correspondence to the rules of the common RLS method.

For the sake of simplicity, this chapter uses solely single channel observations, but we emphasize that the extension to multi-channel signals is most often straightforward as we demonstrate in some of the supplementary examples.

## 4.1 Autonomous Linear State Space Models (ALSSMs)

The output  $y_k \in \mathbb{R}$  of an autonomous (i.e., input free), time-invariant state space system of order  $N$  is given by

$$\begin{aligned}x_k &= Ax_{k-1} , \\y_k &= cx_k\end{aligned}\tag{4.1}$$

with *state-transition matrix*  $A \in \mathbb{R}^{N \times N}$ , *output vector*  $c \in \mathbb{R}^{1 \times N}$ , *state vector*  $x_k \in \mathbb{R}^{N \times 1}$ , and the sequence index  $k \in \mathbb{Z}$ . We assume here that  $A$  is invertible, giving us more flexibility in the window selection, as we exploit in this chapter. Such an ALSSM generates a wide class of discrete-time signals as a function of the initial state vector  $x_0$  [56, Chapter 3]. Such a signal over index  $k$  is given by

$$s_k(x_0) = cA^k x_0 \in \mathbb{R}.\tag{4.2}$$

Note that ALSSMs do not have an input. Thus, the output sequence  $\{s_k(x_0), k \in \mathbb{Z}\}$  is, for any given initial state  $x_0$ , deterministic. Further, note that for non-invertible  $A$ 's, (4.2) only holds for  $k \geq 0$ .

A list of common ALSSM parametrizations is given in Table 4.1.

**Example 1** (Straight line model using an ALSSM). Assume a linear model such as

$$s_i(x) = a_0 + a_1 i \in \mathbb{R} , \quad (4.3)$$

with known model parameter  $x = [a_0, a_1]^T \in \mathbb{R}^2$ ; to be more concrete, consider the numerical values  $a_0 = -1$  and  $a_1 = 2$ . (In this example,  $a_0$  obviously represents the line offset and  $a_1$  the line slope.) To write (4.3) in a recursive form, we use an ALSSM of order  $N = 2$  and assign

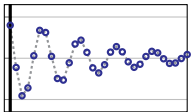
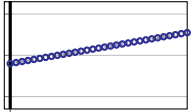
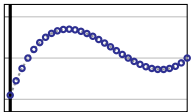
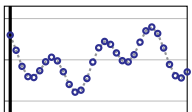
$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} , \quad c = [1, 0] , \quad (4.4)$$

which leads with  $x = [a_0, a_1]^T \in \mathbb{R}^2$  to

$$\begin{aligned} s(x) &= [\dots, s_{-1}(x), s_0(x), s_1(x), s_2(x), s_3(x), \dots] \\ &= [\dots, cA^{-1}x, cA^0x, cA^1x, cA^2x, cA^3x, \dots] \\ &= [\dots, [1, -1]x, [1, 0]x, [1, 1]x, [1, 2]x, [1, 3]x, \dots] \\ &= [\dots, a_0 - a_1, a_0, a_0 + a_1, a_0 + 2a_1, a_0 + 3a_1, \dots] \\ &= [\dots, -3, -1, 1, 3, 5, 7, \dots] . \end{aligned}$$

◇

**Table 4.1:** ALSSM parametrization for a selection of common functions

| <i>Signal <math>s_i(x)</math></i>  | <i>State Space Model <math>s_i(x) = cA^i x</math></i>   | <i>Example</i>   |
|--|---|--|
| <p><b>Sine / Cosine</b><br/> <math>\rho^i [a_0 \cos(\Omega i) - a_1 \sin(\Omega i)]</math><br/> <math>\rho \in \mathbb{R}_+, \Omega \in \mathbb{R}_+</math></p>  | $A = \rho \begin{bmatrix} \cos \Omega & -\sin \Omega \\ \sin \Omega & \cos \Omega \end{bmatrix}$ $c = [1, 0]$ $x = [a_0, a_1]^\top \in \mathbb{R}^{2 \times 1}$   | <br>$i = 0$ |
| <p><b>Line</b><br/> <math>a_0 + a_1 i</math><br/> <math>(a_0, a_1) \in \mathbb{R}^2</math></p>   | $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ $c = [1, 0]$ $x = [a_0, a_1]^\top \in \mathbb{R}^{2 \times 1}$   | <br>$i = 0$ |
| <p><b>Polynomial of order <math>N - 1</math></b><br/> <math>a_0 + a_1 i + \dots + a_{N-1} i^{N-1}</math><br/> <math>(a_0, a_1, \dots, a_{N-1}) \in \mathbb{R}^N</math></p>   | $A = \begin{bmatrix} A_{1,1} & \dots & A_{1,N} \\ \vdots & & \vdots \\ A_{N,1} & \dots & A_{N,N} \end{bmatrix}, A_{n,n'} = \begin{cases} \binom{n'}{n} & \text{if } n \leq n' \\ 0 & \text{if } n > n' \end{cases}$ $c = [1, 0, \dots, 0] \in \mathbb{R}^{1 \times N}$ $x = [a_0, \dots, a_{N-1}]^\top \in \mathbb{R}^{N \times 1}$ | <br>$i = 0$ |
| <p><b>Linear combination of <math>M</math> systems</b><br/> <math>s_i^{(1)}(x^{(1)}) + \dots + s_i^{(M)}(x^{(M)})</math><br/> <math>s_i^{(m)}(x^{(m)}) \in \mathbb{R}</math><br/> <math>m \in \{1, \dots, M\}</math></p> | $A = \text{diag}(A_1, \dots, A_M)$ $c = [c_1, \dots, c_M]$ $x = \left[ (x^{(1)})^\top, \dots, (x^{(M)})^\top \right]^\top$  | <br>$i = 0$ |



## 4.2 ALSSM Model Fitting

Now, we are interested in estimating the initial state of a given ALSSM such that its output sequence fits a given signal around some time index  $k$  by the means of least squared error. To localize the fit “around some time index  $k$ ”, we use a window shifted by  $k$  samples and weight the squared error cost accordingly.

### 4.2.1 Squared Error Cost Functions

Let  $y = [y_1, y_2, \dots, y_K] \in \mathbb{R}^K$  be an observed signal with  $K$  samples. Furthermore, we have an ALSSM of order  $N$  with known parameters  $A \in \mathbb{R}^{N \times N}$  and  $c \in \mathbb{R}^{1 \times N}$ , but unknown initial state  $x_0 \in \mathbb{R}^N$ . Recall that we assume  $A$  to be invertible. Additionally, we have a window  $w$  of finite support; the window writes as  $w = [\dots, w_0, w_1, w_2, \dots]$  with  $w_i \in \mathbb{R}_+$  and  $w_i = 0$  for any  $|i|$  sufficiently large.

We will estimate the initial state  $x_0$  in  $\mathbb{R}^N$  for any index  $k$  such that the localized squared error cost function is minimized. Note that we get a solution for each  $k$ , subsequently denoted as  $\hat{x}_k$ . The estimation for  $\hat{x}_k$  writes as

$$\hat{x}_k = \operatorname{argmin}_{x \in \mathbb{R}^N} \sum_{i=-\infty}^{+\infty} w_{i-k} (cA^{i-k}x - y_i)^2, \quad (4.5)$$

assuming  $y_i = 0$  wherever the index  $i$  exceeds the boundaries of the observations of  $K$  samples. To minimize (4.5), we first reparametrize its cost term as

$$J_k(x) = \sum_{i=-\infty}^{+\infty} w_{i-k} (cA^{i-k}x - y_i)^2 = x^\top W_k x - 2x^\top \xi_k + \kappa_k \quad (4.6)$$

with

$$W_k \triangleq \sum_{i=-\infty}^{+\infty} w_{i-k} (A^{i-k})^\top c^\top c A^{i-k} \in \mathbb{R}^{N \times N}, \quad (4.7)$$

$$\xi_k \triangleq \sum_{i=-\infty}^{+\infty} w_{i-k} (A^{i-k})^\top c^\top y_i \in \mathbb{R}^{N \times 1}, \quad (4.8)$$

$$\kappa_k \triangleq \sum_{i=-\infty}^{+\infty} w_{i-k} y_i^2 \quad \text{in } \mathbb{R}. \quad (4.9)$$

A derivation in detail for this reparametrization is given in (A.1). We note that the right-hand side term in (4.6) is a quadratic form in  $x$  and is, thus, minimized for any invertible  $W_k$  by

$$\hat{x}_k = W_k^{-1} \xi_k . \quad (4.10)$$

A proof for this last step is given, in a more generalized form, in Section 4.2.2.

## 4.2.2 Model Fitting With Linear Constraints

The initial state vector  $x$  is often subject to constraints in the form of linear dependencies expressed as

$$x = Hv + h , \quad (4.11)$$

with fixed  $H \in \mathbb{R}^{N \times M}$  and  $h \in \mathbb{R}^N$ , and remaining unknown  $v \in \mathbb{R}^M$ ,  $M \leq N$ . Imposing these constraints on  $x$ , the minimization of the right-hand side of (4.6) as in (4.10) modifies to

$$\hat{v}_k = (H^T W_k H)^{-1} H^T (\xi_k - W_k h) \quad (4.12)$$

with  $(H^T W_k H)$  assumed invertible.

*Proof.* We substitute  $x$  in (4.6) by (4.11), i.e.,

$$\begin{aligned} \hat{v} &= \underset{v \in \mathbb{R}^M}{\operatorname{argmin}} (v^T H^T W_k H v - 2v^T H^T (\xi_k - W_k h)) \\ &= (H^T W_k H)^{-1} H^T (\xi_k - W_k h) , \end{aligned} \quad (4.13)$$

where in the last step the derivative with respect to  $v$  is computed and set to zero.  $\square$

Note that (4.6) is of a quadratic form in  $x$  and, thus, is minimized in closed form according to (4.13). If constraints on  $x \in \mathbb{R}^N$  apply, i.e., (4.6) is minimized for a subset  $x \subset \mathbb{R}^N$ , then a wide field of other closed form solutions or numerical optimization methods are known [56, Chapter 6].

**Example 2** (Constrained local fit with sinusoidal signal). We locally fit a sinusoidal signal of discrete-time frequency  $\omega$  and of known phase  $\phi \in [-\pi, \pi]$  to the observations  $y$ ; the fit is done locally around index  $k_0$ , using a rectangular window of  $L$  samples. Note that the only unknown parameter in this task is the signal's amplitude. Thus, we use an ALSSM with

$$A = \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix}, \quad c = [1, 0], \quad (4.14)$$

and constrain the initial state to

$$x = \underbrace{\begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix}}_H v + \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_h \quad (4.15)$$

with the single remaining independent variable  $v \in \mathbb{R}$ , reflecting the signal's amplitude. The amplitude estimate  $\hat{v}$  then immediately follows from (4.12) and, hence,  $\hat{x}$  from (4.11).

◇

### 4.2.3 Recursive Computations and Window ALSSMs

When using a sliding window, the computation of (4.7) to (4.9) needs to be repeated at every index  $k$  and is in general exhaustive. This is where recursion comes in: it massively reduces the computing effort for the terms  $W_k$ ,  $\xi_k$ , and  $\kappa_k$  over an interval of indices.

#### The Window ALSSM

To find a recursive form for (4.6), we first need to find a recursive form for the window  $w$ . Therefore we substitute the window by its own ALSSM of order  $Q$ , subsequently denoted as the *window ALSSM*, with its own parameters  $\check{A} \in \mathbb{R}^{Q \times Q}$ ,  $\check{c} \in \mathbb{R}^{1 \times Q}$ , and  $\check{x}_0 \in \mathbb{R}^{Q \times 1}$  over the interval  $a$  to  $b$ , ( $a \in \mathbb{Z} \cup \{-\infty\}$ ,  $b \in \mathbb{Z} \cup \{+\infty\}$ , and  $a \leq b$ ), i.e.,

$$w_j = \begin{cases} \check{c}\check{A}^j\check{x}_0 & \text{for } a \leq j \leq b, \\ 0 & \text{otherwise,} \end{cases} \quad (4.16)$$

with fixed initial state  $\check{x}_0$ . Note that the parameters for this window ALSSM and the interval borders  $a$  and  $b$  have to be selected carefully to guarantee  $\|w\|_2 < +\infty$ . For the sake of simplicity, we again assume  $\check{A}$  to be invertible.<sup>1</sup>

Applying a window with borders  $a$  and  $b$  as in (4.16), the infinite sum in (4.5) simplifies into to the finite sum

$$\hat{x}_k = \operatorname{argmin}_{x \in \mathbb{R}^N} \sum_{i=k+a}^{k+b} \underbrace{(\check{c}\check{A}^{i-k}\check{x}_0)}_{w_{i-k}} (cA^{i-k}x - y_i)^2. \quad (4.17)$$

---

<sup>1</sup>In many cases invertibility is avoidable if the interval borders and the recursion direction (discussed later in this section) are chosen accordingly.

We equally substitute  $w_k$  in (4.7), (4.8), and (4.9) by (4.16) and get

$$W_k = \sum_{i=k+a}^{k+b} (\check{c}\check{A}^{i-k}\check{x}_0)(A^{i-k})^\top c^\top c A^{i-k} \in \mathbb{R}^{N \times N}, \quad (4.18)$$

$$\xi_k = \sum_{i=k+a}^{k+b} (\check{c}\check{A}^{i-k}\check{x}_0)(A^{i-k})^\top c^\top y_i \in \mathbb{R}^{N \times 1}, \quad (4.19)$$

$$\kappa_k = \sum_{i=k+a}^{k+b} (\check{c}\check{A}^{i-k}\check{x}_0)y_i^2 \in \mathbb{R}. \quad (4.20)$$

At this point, we need to introduce the *element-wise* product of two ALSSMs as already proposed in [55].

**Proposition 1** (Product of ALSSM signals [55]). *Let  $s_j^{(1)}$  and  $s_j^{(2)}$ ,  $j \in \mathbb{Z}$ , be two ALSSM signals with respective parameters  $\{c_1, A_1, x_1\}$  and  $\{c_2, A_2, x_2\}$ ,  $A_1$  and  $A_2$  invertible. Then,  $s_j^{(1)} \cdot s_j^{(2)}$ ,  $j \in \mathbb{Z}$ , is also a ALSSM signal with parameters  $\{c_1 \otimes c_2, A_1 \otimes A_2, x_1 \otimes x_2\}$ , where  $\otimes$  denotes the Kronecker product.*

*Proof of Proposition 1.* For any  $j \in \mathbb{Z}$ , we have

$$\begin{aligned} s_j^{(1)} \cdot s_j^{(2)} &= (c_1 A_1^j x_1)(c_2 A_2^j x_2) \\ &= (c_1 A_1^j x_1) \otimes (c_2 A_2^j x_2) \\ &= (c_1 \otimes c_2)(A_1 \otimes A_2)^j (x_1 \otimes x_2). \end{aligned} \quad (4.21)$$

□

Applying Proposition 1, we assign

$$\widehat{c} \triangleq c \otimes \check{c}, \quad (4.22)$$

$$\widehat{A} \triangleq A \otimes \check{A}, \quad (4.23)$$

$$\widehat{x} \triangleq x \otimes \check{x}_0 \quad (4.24)$$

and rewrite (4.18) to (4.20) as

$$W_k = \widehat{W}_k R, \quad (4.25)$$

$$\xi_k^\top = \widehat{\xi}_k^\top R, \quad (4.26)$$

$$\kappa_k = \widehat{\kappa}_k \check{x}_0 \quad (4.27)$$

with

$$\widehat{W}_k = \sum_{i=k+a}^{k+b} (A^{i-k})^\top c^\top \widehat{c} \widehat{A}^{i-k} \in \mathbb{R}^{N \times (NQ)}, \quad (4.28)$$

$$\widehat{\xi}_k^\top = \sum_{i=k+a}^{k+b} \widehat{c} \widehat{A}^{i-k} y_i \in \mathbb{R}^{1 \times (NQ)}, \quad (4.29)$$

$$\widehat{\kappa}_k = \sum_{i=k+a}^{k+b} \check{c} \check{A}^{i-k} y_i^2 \in \mathbb{R}^{1 \times Q}, \quad (4.30)$$

and

$$R = (I_N \otimes \check{x}_0) \in \mathbb{R}^{(NQ) \times N}, \quad (4.31)$$

with full derivations given in (A.6), (A.7), and (A.8). Note that  $W_k \in \mathbb{R}^{N \times N}$  is a square matrix, however  $\widehat{W}_k \in \mathbb{R}^{N \times (NQ)}$  is not.

### The Recursions

While there are in general no immediate recursive forms for  $W_k$ ,  $\xi_k$ , or  $\kappa_k$ , there are recursions for  $\widehat{W}_k$ ,  $\widehat{\xi}_k$ , and  $\widehat{\kappa}_k$ : the forward recursions  $k \rightarrow k+1$  are

$$\widehat{W}_{k+1} = A^{-\top} \widehat{W}_k \widehat{A}^{-1} + C \quad (4.32)$$

$$\text{with } C = -(A^{a-1})^\top c^\top \widehat{c} \widehat{A}^{a-1} + (A^b)^\top c^\top \widehat{c} \widehat{A}^b,$$

$$\widehat{\xi}_{k+1}^\top = \widehat{\xi}_k^\top \widehat{A}^{-1} - \widehat{c} \widehat{A}^{a-1} y_{k+a} + \widehat{c} \widehat{A}^b y_{k+b+1}, \quad (4.33)$$

$$\widehat{\kappa}_{k+1} = \widehat{\kappa}_k \check{A}^{-1} - \check{c} \check{A}^{a-1} y_{k+a}^2 + \check{c} \check{A}^b y_{k+b+1}^2, \quad (4.34)$$

and the backward recursions  $k \rightarrow k-1$  are

$$\widehat{W}_{k-1} = A^\top \widehat{W}_k \widehat{A} + C \quad (4.35)$$

$$\text{with } C = (A^a)^\top c^\top \widehat{c} \widehat{A}^a - (A^{b+1})^\top c^\top \widehat{c} \widehat{A}^{b+1}, \quad (4.36)$$

$$\widehat{\xi}_{k-1}^\top = \widehat{\xi}_k^\top \widehat{A} + \widehat{c} \widehat{A}^a y_{k+a-1} - \widehat{c} \widehat{A}^{b+1} y_{k+b}, \quad (4.37)$$

$$\widehat{\kappa}_{k-1} = \widehat{\kappa}_k \check{A} + \check{c} \check{A}^a y_{k+a-1}^2 - \check{c} \check{A}^{b+1} y_{k+b}^2, \quad (4.38)$$

with full derivations given in (A.9) to (A.14). This result is the recursive computation rule for the cost term (4.6),

$$x^\top W_k x - 2x^\top \xi_k + \kappa_k, \quad (4.39)$$

with (4.25), (4.26), and (4.27), and thanks to the window ALSSM (4.16). Note that this quadratic form is still minimized according to (4.10) or, considering linear constraints, according to (4.12).

**Example 3** (Fit of a straight line to a signal using an exponential window). We wish to locally fit, at any index  $k \in \{1, \dots, K\}$ , a line with offset  $a_0 \in \mathbb{R}$  and slope  $a_1 \in \mathbb{R}$  to the observations  $y_1, \dots, y_K$ . The fit is localized by a right-sided exponentially decaying window with decay  $\gamma \in \mathbb{R}$ ,  $0 \ll \gamma < 1$ , i.e.,  $w_{k,i} = \gamma^{i-k}$  for  $i \geq k$  and  $w_{k,i} = 0$  for  $i < k$ .

To find a recursive form for this fitting problem, we write the window  $w_{i,k}$  as a (degenerated) ALSSM according to (4.16) of order  $Q = 1$  with  $\check{A} = [\gamma]$ ,  $\check{c} = [1]$ , and  $\check{x}_0 = [1]$  and with interval borders  $a = 0$  and  $b \rightarrow +\infty$ . The straight line is modeled by another ALSSM with  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ,  $c = [1, 0]$ , and  $x = [a_0, a_1]^\top$ . Then, the localized fit error is according to (4.39), with (4.25) to (4.27),

$$x^\top (\widehat{W}_k R) x - 2x^\top (R^\top \widehat{\xi}_k) + \widehat{\kappa}_k \check{x}_0 \quad (4.40)$$

with  $R = (I_2 \otimes \check{x}_0) = (I_2 \otimes [1]) = I_2$ . The backward recursions for  $\widehat{W}_k$  and  $\widehat{\xi}_k$  are according to (4.35) and (4.37)

$$\widehat{W}_{k-1} = \gamma A^\top \widehat{W}_k A + c^\top \check{c}, \quad (4.41)$$

$$\widehat{\xi}_{k-1}^\top = \gamma \widehat{\xi}_k^\top A + \widehat{c} y_{k-1}, \quad (4.42)$$

$$\widehat{\kappa}_{k-1} = \gamma \widehat{\kappa}_k + y_{k-1}^2. \quad (4.43)$$

Note that we here use the following simplifications:

$$\widehat{A}^a = I_2 \quad \text{for } a=0, \quad (4.44)$$

$$\widehat{A}^{b+1} = (A \otimes [\gamma])^{b+1} = 0 \quad \text{for } b \rightarrow +\infty. \quad (4.45)$$

◇

**Example 4** (Signal fit using a Hamming window). We wish to fit an ALSSM to a signal applying a Hamming window; a Hamming window is known as

$$w_{\ell-k} = 0.54 - 0.46 \cos\left(\frac{2\pi\ell}{L-1}\right), \quad n = 0, 1, \dots, L-1. \quad (4.46)$$

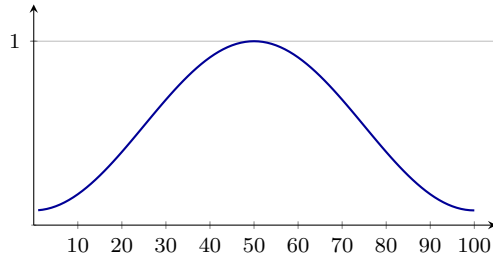
This translates to a window ALSSM with parameters

$$\check{A} = \left[ \begin{array}{cc|c} \cos(2\pi/L) & -\sin(2\pi/L) & 0 \\ \sin(2\pi/L) & \cos(2\pi/L) & 0 \\ 0 & 0 & 1 \end{array} \right], \quad (4.47)$$

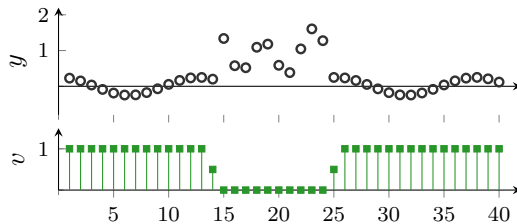
$$\check{c} = [1, 0, 1],$$

$$\check{x}_0 = [-0.46, 0, 0.54]^\top,$$

in the interval  $a = 0$  to  $b = L - 1$ . This window can be applied to any fit as in (4.17). The output of this window ALSSM (4.47) is shown in Fig. 4.1. ◇



**Figure 4.1:** Hamming window of length  $L = 100$  produced using the window ALSSM (4.47).



**Figure 4.2:** Example of signal samples  $y$  (black circles), where each sample comes along with a measure of reliability or *sample weight*  $v \in [0, 1]$  (green stems). The signal  $y$  shown is a discrete-time sinusoidal signal interrupted by a sequence of distorted samples and associated with a sample weight of  $v = 1$  for samples of highest reliability, and with  $v = 0$  for any samples of lowest reliability.

#### 4.2.4 Weighted Samples

So far, we assumed each sample to be of equal importance or weight. However, this does not often sufficiently reflect reality, for example when doing multi-step processing. When signals pass multiple steps, where the output of one preceding step is the input to the subsequent step, intermediate values are often complemented by an individual measure of reliability or uncertainty. Here we should in any subsequent step incorporate this knowledge and value the samples accordingly. To do so, we modify our method and complement the single observations by a specific *sample weight*, as depicted in Fig. 4.2.

Let  $\mathbf{y} = [y_1, y_2, \dots, y_K]^T \in \mathbb{R}^K$  be an observed signal where each sample  $y_i$  is weighted by its associated sample weight  $v_i \in \mathbb{R}_+$ . Thus,

the cost term (4.6) modifies to

$$J_k(x) = \sum_{i=k+a}^{k+b} w_{i-k} v_i (cA^{i-k}x - y_i)^2. \quad (4.48)$$

Accordingly, the forward recursions (4.33) to (4.34) for  $k \rightarrow k+1$  modify to

$$\widehat{W}_{k+1} = A^{-\top} \widehat{W}_k \widehat{A}^{-1} + C(k) \quad (4.49)$$

$$\text{with } C(k) = -v_{k+a} (A^{a-1})^\top c^\top \widehat{c} \widehat{A}^{a-1} + v_{k+b+1} (A^b)^\top c^\top \widehat{c} \widehat{A}^b,$$

$$\widehat{\xi}_{k+1}^\top = \widehat{\xi}_k^\top \widehat{A}^{-1} - v_{k+a} \widehat{c} \widehat{A}^{a-1} y_{k+a} + v_{k+b+1} \widehat{c} \widehat{A}^b y_{k+b+1}, \quad (4.50)$$

$$\widehat{\kappa}_{k+1} = \widehat{\kappa}_k \widehat{A}^{-1} - v_{k+a} \check{c} \check{A}^{a-1} y_{k+a}^2 + v_{k+b+1} \check{c} \check{A}^b y_{k+b+1}^2, \quad (4.51)$$

and the backward recursions (4.35) to (4.38) for  $k \rightarrow k-1$  to

$$\widehat{W}_{k-1} = A^\top \widehat{W}_k \widehat{A} + C(k) \quad (4.52)$$

$$\text{with } C(k) = v_{k+a-1} (A^a)^\top c^\top \widehat{c} \widehat{A}^a - v_{k+b} (A^{b+1})^\top c^\top \widehat{c} \widehat{A}^{b+1},$$

$$\widehat{\xi}_{k-1}^\top = \widehat{\xi}_k^\top \widehat{A} + v_{k+a-1} \widehat{c} \widehat{A}^a y_{k+a-1} - v_{k+b} \widehat{c} \widehat{A}^{b+1} y_{k+b}, \quad (4.53)$$

$$\widehat{\kappa}_{k-1} = \widehat{\kappa}_k \widehat{A} + v_{k+a-1} \check{c} \check{A}^a y_{k+a-1}^2 - v_{k+b} \check{c} \check{A}^{b+1} y_{k+b}^2. \quad (4.54)$$

**Example 5** (A non-equidistant sampling scheme). Let  $k \in \{1, \dots, K\}$ ,  $K \in \mathbb{N}$ , be an equidistant time index. Now, we assume an ADC device which only takes samples at some time indices  $k$  but not at the others according to a non-equidistant sampling scheme; whenever the ADC takes a sample, we denote the sample  $y_k \in \mathbb{R}$  and set its associated sample weight to  $v_k = 1$ . For all other samples, we set  $y_k = 0$  and  $v_k = 0$ . Then we directly apply (4.48) to fit our models to this non-equidistantly sampled signal.

◇

Note that, for any sparse  $v_k$ , most of the sum terms in (4.49) to (4.51) and in (4.52) to (4.54) are zero (i.e., multiplied by  $v_k = 0$ ) and therefore the recursion simplifies. For example, the term  $A^{-\top} \widehat{W}_k \widehat{A}^{-1}$  in (4.49) simplifies for  $M$  iteration steps with zero weighted observations to

$$\underbrace{(A^{-\top} \dots (A^{-\top} \widehat{W}_k \widehat{A}^{-1}) \dots \widehat{A}^{-1})}_{M \text{ times}} = (A^{-M})^\top \widehat{W}_k (\widehat{A}^{-M}). \quad (4.55)$$

## 4.2.5 Time-Variant Windows

In some cases, we even have recursions if the window ALSSM is time-dependent, or more precisely, when each sample comes along with its



own transition matrix  $\check{A}_i$  for the window ALSSM. In the following, we look at the case of a first order system, since this simple case has an immediate practical application as is already shown in [47].

Let the parameters of the window ALSSM be  $\check{A}_i = [\gamma_i] \in \mathbb{R}$ ,  $\check{c} = [1]$ , leading to the time-variant system<sup>2</sup>

$$\begin{aligned} \check{x}_j &= \gamma_{k+j} \check{x}_{j-1} , \\ \check{w}_{k,j} &= \check{c} \check{x}_j . \end{aligned} \quad (4.56)$$

Using this time-variant window, (4.17) modifies to

$$\hat{x}_k = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \sum_{i=k+a}^{k+b} v_i \underbrace{(\check{c} \alpha_k(i) \check{x}_0)}_{w_{k,i-k}} (cA^{i-k}x - y_i)^2 \quad (4.57)$$

with

$$w_{k,j} = \begin{cases} \check{c} \alpha_k(k+j) \check{x}_0 & \text{for } a \leq j \leq b , \\ 0 & \text{otherwise ,} \end{cases} \quad (4.58)$$

and with  $\alpha_k(i) \in \mathbb{R}$  defined as in [47],

$$\alpha_k(i) \triangleq \begin{cases} \prod_{j=i+1}^k \gamma_j^{-1} & \text{if } k > i , \\ \prod_{j=k+1}^i \gamma_j & \text{if } k < i , \\ 1 & \text{if } k = i . \end{cases} \quad (4.59)$$

Summing the weights of the samples involved in the cost term (4.57) after applying the cost's window, leads to the value of *weighted number of samples* and is defined as

$$\nu_k \triangleq \sum_{i=k+a}^{k+b} v_i w_{k,i-k} \in \mathbb{R}_+ . \quad (4.60)$$

We here notice that this weighted number of samples varies with the sample weights  $v_i$ ; it is not guaranteed, that there is a minimum number of samples involved in a cost term if the samples within the window are weighted unfavorable.

If we want to control  $\nu_k$ , i.e., to have  $\nu_k$  fixed and independent of  $k$  and also independent of any sample weights, i.e., we want  $\nu_k = g$ ,

---

<sup>2</sup>Note that we use index  $k$  as the evaluation index, i.e., the localization index of the window;  $i$  denotes the absolute index of a sample along the signal; and index  $j$  is the relative index of a sample within the window with  $j = i - k$ .

$g \geq 0$  with a fixed  $g \in \mathbb{R}_+$  for any  $k$ , the window needs to adapt locally depending on the sample weights  $v_i$ ; the window needs to get wider wherever the sample weights  $v_i$  are small, and to get smaller wherever the  $v_i$ 's are large.

This behavior is achieved with the following trick: for a right-side window, i.e.,  $a = 0$  and  $b \rightarrow +\infty$ , we have

$$\nu_k = \sum_{i=k}^{+\infty} v_i \underbrace{(\check{c}\alpha_k(i)\check{x}_0)}_{w_{k,i-k}} \quad (4.61)$$

with the recursions

$$\begin{aligned} \nu_{k-1} &= \sum_{i=k-1}^{+\infty} v_i (\check{c}\alpha_{k-1}(i)\check{x}_0) \\ &= \sum_{i=k}^{+\infty} v_i (\check{c}\alpha_{k-1}(i)\check{x}_0) + v_{k-1} \underbrace{(\check{c}\alpha_{k-1}(k-1)\check{x}_0)}_1 \\ &= \gamma_k \sum_{i=k}^{+\infty} v_i (\check{c}\alpha_k(i)\check{x}_0) + v_{k-1} \check{c}\check{x}_0 \\ &= \gamma_k \nu_k + v_{k-1} \check{c}\check{x}_0, \end{aligned} \quad (4.62)$$

where we use for the second last step the relation  $\alpha_{k-1}(i) = \gamma_k \alpha_k(i)$ . Then, if we force  $\nu_k = \nu_{k-1} = g$  for any indices  $k$ , we get for this right-sided window

$$\overrightarrow{\gamma}_k = \frac{g - v_{k-1} \check{c}\check{x}_0}{g}. \quad (4.63)$$

Correspondingly, we get for a left-sided window

$$\overleftarrow{\gamma}_k = \frac{g}{g - v_k \check{c}\check{x}_0}. \quad (4.64)$$

An example using this mechanism is given later in Section 9 on page 99.

## 4.2.6 Multi-Segment Windows

We have discussed in Section 4.2.3 the use of single windows over a finite or infinite interval. To generate more versatile windows, multiple such windows can be joined, where each window comes along with its own

window ALSSM. Thereby, the intervals do not necessarily need to be directly adjacent; gaps in between the individual intervals may remain.

We assign to the  $p$ th interval the ALSSMs of parameters  $A^{(p)}$ ,  $c^{(p)}$ , and common  $x$ , and the window ALSSM of parameters  $\check{A}^{(p)}$ ,  $\check{c}^{(p)}$ , and common  $\check{x}$ . Then, each segment comes along with its own cost function according to (4.39),

$$x^\top W_k^{(p)} x - 2x^\top \xi_k^{(p)} + \kappa_k^{(p)}. \quad (4.65)$$

It follows that the total cost function over all segments is

$$\begin{aligned} \tilde{J}_k(x) &= \sum_p \left( x^\top W_k^{(p)} x - 2x^\top \xi_k^{(p)} + \kappa_k^{(p)} \right) \\ &= x^\top \underbrace{\sum_p \left( W_k^{(p)} \right)}_{\tilde{W}_k} x - 2x^\top \underbrace{\sum_p \left( \xi_k^{(p)} \right)}_{\tilde{\xi}_k} + \underbrace{\sum_p \left( \kappa_k^{(p)} \right)}_{\tilde{\kappa}_k} \\ &= x^\top \tilde{W}_k x - 2x^\top \tilde{\xi}_k + \tilde{\kappa}_k \end{aligned} \quad (4.66)$$

with

$$\tilde{W}_k = \sum_p \left( W_k^{(p)} \right), \quad \tilde{\xi}_k = \sum_p \left( \xi_k^{(p)} \right), \quad \tilde{\kappa}_k = \sum_p \left( \kappa_k^{(p)} \right). \quad (4.67)$$

Note that all segments use a common state vector  $x$ . This is not a loss in generality since multiple ALSSM systems might be stacked using block diagonals (cf. Table 4.1). We note that (4.66) is again of the same quadratic form as (4.6), and therefore also minimized by (4.10).

## 4.3 The Hilbert Space Perspective

### 4.3.1 The Localized Hilbert Space $\mathcal{H}$

A Hilbert space [45] is essentially a vector space with an inner product.<sup>3</sup> We here denote a *localized* Hilbert space as a space defined with a localized inner product.

Let  $y = [y_1, y_2, \dots, y_K] \in \mathbb{R}^K$  and  $y' = [y'_1, y'_2, \dots, y'_K] \in \mathbb{R}^K$  be elements in a localized Hilbert space  $\mathcal{H}$  spanned by  $\mathbb{R}^K$ . Furthermore,

<sup>3</sup>A Hilbert space requires in addition to the inner product the convergence of every Cauchy sequence in the space. This is usually true for practical applications and, thus, here not further discussed.

we have the weights  $w = [w_1, w_2, \dots, w_K] \in \mathbb{R}_+^K$  as in (4.5). Then the (localized) inner product in  $\mathcal{H}$  is defined as

$$\begin{aligned} \langle y, y' \rangle_w &\triangleq \sum_{i=1}^K w_i y_i y'_i \\ &= y^\top \text{diag}(w) y' \end{aligned} \quad (4.68)$$

and, hence, the (localized) norm writes as

$$\|y\|_w \triangleq \sqrt{\langle y, y \rangle_w} . \quad (4.69)$$

### 4.3.2 The Localized Feature Vector Space $\mathcal{F}$ — Another Hilbert Space

With prior knowledge of the signal structure of  $y$ , the norm in (4.69) can often be computed more efficiently. In particular, we here consider the signals  $y \in \mathbb{R}^K$  of the form  $y_i = cA^i x$  with  $y_i$  the  $i$ th element of  $y$ , and an ALSSM with invertible  $A \in \mathbb{R}^{N \times N}$ , and with  $c \in \mathbb{R}^{1 \times N}$  and  $x \in \mathbb{R}^N$ . This is a linear map of the form

$$\mathbb{R}^N \rightarrow \mathbb{R}^K : x \mapsto y \quad (4.70)$$

with the feature vector  $x$  in a feature vector space in  $\mathbb{R}^N$ .

Supplementing this feature vector space with its own inner product

$$\begin{aligned} \langle x, x' \rangle_{\mathcal{F}, w} &\triangleq \sum_{i=1}^K w_i y_i y'_i \\ &= \sum_{i=1}^K w_i \overbrace{(cA^i x)}^{y_i} \overbrace{(cA^i x')}^{y'_i} \\ &= x^\top \underbrace{\left[ \sum_{i=1}^K w_i ((A^i)^\top c^\top) (cA^i) \right]}_W x' \\ &= x^\top W x' \end{aligned} \quad (4.71)$$

with a window  $w \in \mathbb{R}^K$ , leads to a new Hilbert space denoted as  $\mathcal{F}$ . Hence, the norm of  $\mathcal{F}$  writes as

$$\|x\|_{\mathcal{F}, w} \triangleq \sqrt{\langle x, x \rangle_{\mathcal{F}, w}} . \quad (4.72)$$

Note that the size of the feature vector  $x$  is independent of the size of  $y$ , and we commonly choose  $N \ll K$ . Therefore, it is often preferred to use  $x$  as a lower dimensional *surrogate* of  $y$  and to compute the norm using (4.72) instead of (4.69).

### 4.3.3 An Euclidean Feature Vector Space

Similarly, under some linear transformations, as for the ALSSMs, the norm (4.72) can be manipulated and often simplified. Applying a linear transformation to the feature vector space  $\mathcal{F}$  from Section 4.3.2 turns it into an Euclidean feature vector space, with an Euclidean norm and inner product. This transformation often simplifies any further processing.

Starting from (4.71), we write

$$\begin{aligned} \langle x, x' \rangle_{\mathcal{F}, w} &= x^{\top} W x' \\ &= (x^{\top} V^{\top})(V x') \\ &= z^{\top} z' \\ &= \langle z, z' \rangle, \end{aligned} \tag{4.73}$$

where we use in the second step the Cholesky decomposition,

$$W = V^{\top} V \tag{4.74}$$

of the positive semi-definite matrix  $W$  and substitute  $z = Vx$  and  $z' = Vx'$ . This substitution is the linear transformation

$$\mathbb{R}^N \rightarrow \mathbb{R}^N : x \mapsto Vx, \tag{4.75}$$

mapping elements from the feature vector space  $\mathcal{F}$  to a transformed feature vector space with an Euclidean inner product and norm. An application to this mapping is given later on in Section 5.3.

### 4.3.4 The Feature Vectors as a Signal Surrogate

Summarizing Sections 4.3.2 and 4.3.3, we have for any two signals  $y \in \mathcal{F}$  and  $y' \in \mathcal{F}$  with corresponding  $x \in \mathbb{R}^N$  and  $x' \in \mathbb{R}^N$ , and  $z \in \mathbb{R}^N$  and  $z' \in \mathbb{R}^N$ , the following interesting relations:

$$\langle y, y' \rangle_w = \langle x, x' \rangle_{\mathcal{F}, w} = \langle z, z' \rangle, \tag{4.76}$$

$$\|y\|_w = \|x\|_{\mathcal{F}, w} = \|z\|, \tag{4.77}$$

where the rightmost terms are Euclidean inner products and norms.

Recall that  $y$  is of length  $K$  and is expressed by the output sequence of an ALSSM of order  $N$ , with fixed parameters  $A$  and  $c$ , but with free initial state  $x$ .  $\mathcal{F}$  is a Hilbert space spanned by the initial states  $x \in \mathbb{R}^N$  with a non-Euclidean inner product (see Section 4.3.2), and  $z$  is a transformed version of  $x$  with an Euclidean inner product (see Section 4.3.3). As we commonly choose  $N \ll K$ , it is often beneficial to use for any further processing the states  $z$  (or also  $x$ ) as a surrogate instead of the much larger time signal  $y$ .

As a final remark, we note that  $\mathcal{F}$  is also a Hilbert space if  $x$  is an element of a subspace of  $\mathbb{R}^N$ , e.g., also if linear constraints are applied on  $x$  as in Section 4.2.2.

## 4.4 The Feature-Vector Perspective

The unconstrained minimization (4.5) for a signal  $y \in \mathbb{R}^K$  at any index  $k \in \{1, \dots, K\}$  is a linear mapping in  $y$ ,

$$\mathbb{R}^K \times \mathbb{Z} \rightarrow \mathbb{R}^N : (y, k) \mapsto \hat{x}_k . \quad (4.78)$$

Applying this mapping at every index  $k$  results in a new *feature vector signal*  $\hat{x} = [\hat{x}_1, \dots, \hat{x}_K] \in \mathbb{R}^{N \times K}$ . This immediately leads to the ALSSM filtering application in Section 4.5 and is the basis for the clustering application later in Section 5.5.

## 4.5 ALSSMs as LTI Filters

Choosing a selection vector  $c_0 \in \mathbb{R}^{1 \times N}$  and applying the mapping (4.78) leads to a linear time-invariant (LTI) single-input single-output filter

$$\hat{y}_k = c_0 \hat{x}_k . \quad (4.79)$$

This application is closely related to Savitzky-Golay filters [41]. Note that the extension to  $M$  inputs or/and  $L$  outputs is straightforward by extending  $c \in \mathbb{R}^{M \times N}$  or/and  $c_0 \in \mathbb{R}^{L \times N}$ .

**Example 6** (Signal smoothing with a one-sided or two-sided polynomial ALSSM filter). Let  $y \in \mathbb{R}^K$  be a recorded signal. To get a smoothed version of that signal, we apply an ALSSM filter using a second order polynomial model with

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \quad c = [1, 0, 0],$$

supplemented with a left-sided exponential decaying window with decay factor  $\gamma \in \mathbb{R}_+$ ,  $0 \ll \gamma < 1$ , i.e., the window ALSSM writes as

$$\check{A} = [\gamma^{-1}], \quad \check{c} = [1], \quad \check{x}_0 = [1]^\top,$$

with interval borders  $a \rightarrow -\infty$  and  $b = -1$ . Note that this left-sided window ALSSM needs to be computed in forward direction  $k \rightarrow k + 1$  to get stable recursions. The output selection vector for this filter is chosen as  $c_0 = c$ . An example of this one-sided, non-symmetric filter's output is shown Fig. 4.3.

If a symmetric filter is needed, the window needs to be extended to a two-sided and symmetric window: we add a second cost function with the same ALSSM but with a right-sided window, leading to a multi-segment window as in Section 4.2.6; the corresponding right-sided window ALSSM writes as

$$\check{A} = [\gamma], \quad \check{c} = [1], \quad \check{x}_0 = [1]^\top,$$

with interval borders  $a = 0$  and  $b \rightarrow +\infty$ . Note that this right-sided window ALSSM computes in backward direction  $k \rightarrow k - 1$ . The output of this symmetric filter  $\hat{y}'$  with this two-sided multi-segment window is also shown in Fig. 4.3.

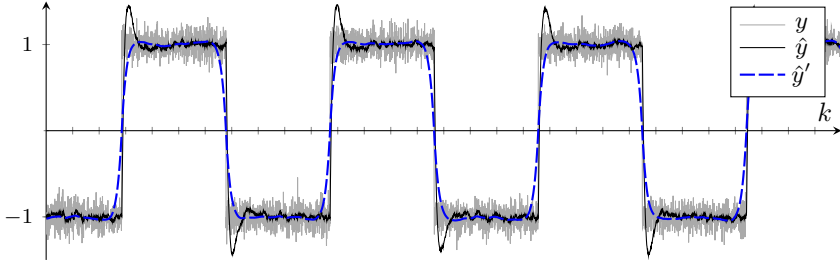
◇

## 4.6 Localization of Events Using Local Cost Ratios (CRs)

Often we are interested in localizing certain events in a signal. Such events include pulses of particular shapes, changes in the signal's characteristics, and outlier samples or outlier sequences.<sup>4</sup> All these events have in common that the accuracy with which a selected model fits on the signal abruptly changes. A good measure for accuracy is a *cost ratio* (CR) or *log-cost ratio* (LCR), which compare the fit of two alternative models; both ratios peak at event locations.

---

<sup>4</sup>We here refer to outliers as samples or sequences that have signal qualities that differ from the rest of the signal.



**Figure 4.3:** ALSSM filter as in Example 6 with output  $\hat{y}_k = c_0 x_k$ . Input signal  $y$  (gray); filter output  $\hat{y}$  (black line) applying a one-sided, non-symmetric model; alternative filter output  $\hat{y}'$  using a two-sided, symmetric model (dashed blue line).

CRs or LCRs correspond to *likelihood ratios* (LRs) or *log-likelihood ratios* (LLRs) as known from statistical theory; LRs and LLRs are the probability ratios between two hypotheses. At our institute, localized versions of model likelihoods were derived to detect certain events such as model switches or pulses in given observations [29], [56, Section 7.3], [39, Part II]; [38] applies the likelihood to estimate a model with constrained parameters. In this thesis, CRs and LCRs are used in various applications; a prominent example follows later on in Chapter 7, localizing the cardiac activation.

Let  $J_k(x) \in \mathbb{R}$  be the squared error of an ALSSM parametrized by its state vector  $x$  and fitting a signal  $y$  around time index  $k$ , if an event occurs at index  $k$ .  $J'_k(x') \in \mathbb{R}$  is an alternative model parametrized by its own state vector  $x'$  and fitting the same signal around the same index  $k$ , if there is no event occurrence at index  $k$ . Then we define the local cost ratio as

$$\text{CR}_k \triangleq \frac{\min_x J_k(x)}{\min_{x'} J'_k(x')} \in \mathbb{R}_+ \quad (4.80)$$

which peaks at every occurrences of the event [39, Section 6].

In statistical theory, this ratio is, up to irrelevant constants, interpreted as a likelihood ratio between two hypotheses [56]. In analogy, we



additionally introduce the local log-cost ratio

$$\text{LCR}_k \triangleq -\frac{1}{2} \log (\text{CR}_k) = -\frac{1}{2} \log \frac{\min_x J_k(x)}{\min_{x'} J'_k(x')} \in \mathbb{R} . \quad (4.81)$$

This logarithmic measure, the LCR, preserves the local minima and maxima of the CR (up to a sign switch) and is often more practical to be displayed.

### 4.6.1 Detection of Pulses

When detecting a pulse or a class of pulses within a signal, we need an ALSSM approximating the particular shape(s). Let  $J_k(x)$  be the error when fitting that ALSSM to the signal at index  $k$ , assuming the pulse to be located at index  $k$ ;  $J_k(\mathbf{0})$  is the error when we assume to have no pulse. Good candidates for pulse locations are then where the LCR as in (4.80) is locally maximal, i.e., where

$$\text{LCR}_k = -\frac{1}{2} \log \frac{\min_x J_k(x)}{J_k(\mathbf{0})} \in \mathbb{R} \quad (4.82)$$

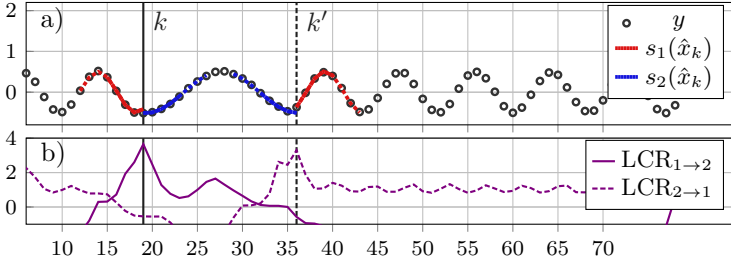
peaks.

This method is used later on in Example 3, Section 5.5, and extensively in [47, 48]. Further references to more examples are given in Section 10.1.

### 4.6.2 Detection of Model Switches

A model switch denotes the change in the signal generator of an observed signal. Model switches are, for example, observed at the starting point and the ending point of a pulse in a recorded signal, or at the transition from one signal form to another as in Fig. 4.4.

Let  $y$  be the observations of a signal generator switching between two different models. Further,  $\overleftarrow{J}_k^{(1)}(x_1)$  denotes the error when fitting an ALSSM corresponding to the first model using a left-sided window (i.e.,  $w_k = 0$  for  $k \geq 0$ ), and  $\overrightarrow{J}_k^{(1)}(x_1)$  using a right-sided window (i.e.,  $w_k = 0$  for  $k < 0$ ).  $\overleftarrow{J}_k^{(2)}(x_2)$  and  $\overrightarrow{J}_k^{(2)}(x_2)$  denote the errors fitting the second model with its own state  $x_2$  and corresponding windows.



**Figure 4.4:** Toy example of model switch detection using log-cost ratios (LCR) as in Section 4.6.2. The upper plot (a) shows: observations  $y$  (black circles) generated by two alternating sinusoidal models of different frequencies; model trajectories at detected model switches from model 1 ( $s_1(\hat{x}_k)$ , red line), to model 2 ( $s_2(\hat{x})$ , blue line) at index  $k = 19$ , and from model 2 to model 1 at index  $k' = 36$ . The lower plot (b) shows: log-cost ratio  $LCR_{1 \rightarrow 2}$  (solid line) with local maxima.

Then we find the switches from the first to the second model at the local maxima of

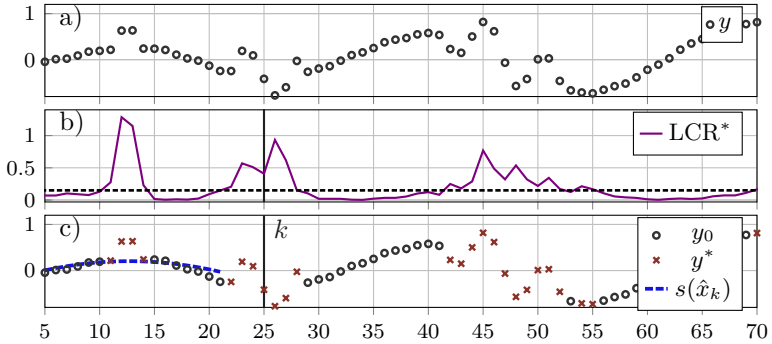
$$LCR_k = -\frac{1}{2} \log \frac{\min_x \overleftarrow{J}_k^{(1)}(x) + \min_x \overrightarrow{J}_k^{(2)}(x)}{\min_x \left( \overleftarrow{J}_k^{(1)}(x) + \overrightarrow{J}_k^{(1)}(x) \right)} \in \mathbb{R}. \quad (4.83)$$

It's vice versa to detect switches from the second model back to the first model. A toy example using this method to detect model switches is given in Fig. 4.4. Furthermore, model switches are also discussed, in more detail, in [47, Section III.G] and in [56].

### 4.6.3 Detection of Outliers

We refer to outliers in a signal as samples or sequences whose properties differ in one way or another from those of the other samples, as for example in Fig. 4.5.

Let  $y$  be the signal of a signal generator, interrupted by some outlier samples. Furthermore,  $J(x)$  is the error function locally fitting the (outlier free) signal appropriately well;  $J'(x)$  is the error function using the same model and the same window with the only difference being that its window is set to zero over a short interval, i.e.,  $w_i = 0$  for  $i \in \{a, \dots, b\}$ , for  $a < b$ ,  $a \in \mathbb{Z} \cup \{-\infty\}$ ,  $b \in \mathbb{Z} \cup \{+\infty\}$ .



**Figure 4.5:** Toy example of outlier detection according to Section 4.6.3, using an ALSSM with a second order polynomial model and a two-sided exponentially decaying window with a gap interval of 5 samples. Plot (a): observations  $y$  with outliers. Plot (b): log-cost ratio LCR (solid line) with threshold for outlier detection (dashed line). Plot (c): detected outliers  $y^*$ , remaining observations  $y_0$ , and local model fit of model  $s(\hat{x}_k)$  with state estimate  $x_k$  localized at index  $k = 13$  (red dashed line).

Then, the local cost ratio

$$\text{LCR}_k = -\frac{1}{2} \log \frac{J'_k(\hat{x})}{J_k(\hat{x})} \in \mathbb{R} \quad (4.84)$$

with

$$\hat{x} = \min_x J'_k(x) \quad (4.85)$$

will rise whenever the model within this gap interval  $\{a, \dots, b\}$  fails and a simple threshold on the LCR will do the outlier discrimination. Such an example is given in Fig. 4.5.

Note that outlier detection using our ALSSM method is also discussed in more detail in [47, Section III.G].



# Chapter 5

## Model-to-Model Fit Using ALSSMs

In Chapter 4, we have fit ALSSM output sequences to discrete-time signals. In this chapter, we compare the output sequences of two ALSSMs using the squared error cost, and we minimize this cost by optimizing the initial state of one or both ALSSMs (while leaving model parameters  $A$  and  $c$  fixed).

We focus on optimizations in the feature space of ALSSMs because complex optimization problems can often be broken into simpler successive optimization steps, whereby each step only relies on the output, i.e., the feature vectors, of the previous step.

In our examples, we commonly extract a feature vector as a first step by applying an ALSSM fit. The later optimization steps map this extracted feature vector onto more sophisticated feature vectors. While this multi-step approach generally entails the risk of error accumulation, this risk is avoidable if the individual steps are carefully selected, as we will show later.

### 5.1 Fit of Any Two ALSSMs

Given are two ALSSMs, the first of order  $N_1 \in \mathbb{N}$  with fixed parameters  $A_1$  and  $c_1$  and output sequence  $y_1$  with elements  $y_{1,k} = c_1 A_1^k x_1$ ; the second of order  $N_2 \in \mathbb{N}$  with fixed parameters  $A_2$  and  $c_2$  and output sequence  $y_2$  with elements  $y_{2,k} = c_2 A_2^k x_2$ . We are interested in optimizing the

initial states  $x_1 \in \mathbb{R}^{N_1}$  and/or  $x_2 \in \mathbb{R}^{N_2}$  of the two ALSSMs such that the localized squared error  $\|y_1 - y_2\|_w^2$  as in (4.69) is minimum. The localization is again expressed by the window  $w = [\dots, w_0, w_1, w_2, \dots]$ , with  $w_i \in \mathbb{R}_+$  and  $w_i = 0$  for any  $|i|$  sufficiently large (cf. localized norm in Section 4.3). Therefore, we look at the cost

$$\begin{aligned}
\|y_1 - y_2\|_w^2 &= \sum_{i=-\infty}^{+\infty} w_i \|c_1 A_1^i x_1 - c_2 A_2^i x_2\|^2 \\
&= \sum_{i=-\infty}^{+\infty} w_i (c_1 A_1^i x_1 - c_2 A_2^i x_2)^\top (c_1 A_1^i x_1 - c_2 A_2^i x_2) \\
&= x_1^\top \underbrace{\left( \sum_{i=-\infty}^{+\infty} w_i (A_1^i)^\top c_1^\top c_1 A_1^i \right)}_{W_{11}} x_1 - 2x_1^\top \underbrace{\left( \sum_{i=-\infty}^{+\infty} w_i (A_1^i)^\top c_1^\top c_2 A_2^i \right)}_{W_{12}} x_2 \\
&\quad + x_2^\top \underbrace{\left( \sum_{i=-\infty}^{+\infty} w_i (A_2^i)^\top c_2^\top c_2 A_2^i \right)}_{W_{22}} x_2 \\
&= x_1^\top W_{11} x_1 - 2x_1^\top W_{12} x_2 + x_2^\top W_{22} x_2 .
\end{aligned} \tag{5.1}$$

For a fixed  $x_2$ , the unconstrained minimization for  $x_1$  follows in analogy to (4.10) as

$$\hat{x}_1 = W_{11}^{-1} W_{12} x_2 \tag{5.2}$$

for any invertible  $W_{11}$  (which is the case for most of the practical model and window selections). The minimization with linear constraints on  $x_1$  is according to Section 4.2.2 after applying the substitutes  $W_k = W_{11}$ ,  $\xi_k = W_{12} x_2$ , and  $\kappa_k = x_2^\top W_{22} x_2$ . For a constrained  $x_2$ , the substitution is analogous.

## 5.2 Fit Using a Window ALSSM

To efficiently compute  $W_{11}$ ,  $W_{12}$ , and  $W_{22}$ , we replace the window  $w$  by its own ALSSM with  $w_i = \check{c} \check{A}^i \check{x}_0$  within bounds  $a \leq i \leq b$  and  $w_i = 0$

otherwise, as in (4.16). Thus, we get using the example of  $W_{11}$ ,

$$\begin{aligned} W_{11} &= \sum_{i=a}^b (\check{c}\check{A}^i\check{x}_0)(A_1^i)^\top c_1^\top c_1 A_1^i \\ &= \widehat{W}_{11}R \end{aligned} \quad (5.3)$$

with

$$\widehat{W}_{11} = \sum_{i=a}^b (A^i)^\top c^\top \widehat{c} \widehat{A}^i, \quad (5.4)$$

and with  $\widehat{c}$ ,  $\widehat{A}$ , and  $R$  as in (4.22), (4.23), and (4.31). The full derivation is in analogy to the derivation in Section 4.2.3.

To simplify computation of  $\widehat{W}_{11}$  for large intervals, we might use the subtraction trick of geometric series,

$$\begin{aligned} A\widehat{W}_{11} - \widehat{W}_{11}\widehat{A}^{-1} &= \sum_{i=a}^b (A^{i+1})^\top c^\top \widehat{c} \widehat{A}^i - \sum_{i=a-1}^{b-1} (A^{i+1})^\top c^\top \widehat{c} \widehat{A}^i \\ &= (A^{b+1})^\top c^\top \widehat{c} \widehat{A}^b - (A^a)^\top c^\top \widehat{c} \widehat{A}^{a-1} \end{aligned} \quad (5.5)$$

which is then a standard problem of the form “ $AX - XB = C$ ” and solved for  $\widehat{W}_{11}$  by solving a linear equation system [2]. For  $W_{12}$  and  $W_{22}$  we proceed similarly.

### 5.3 Distance Between Two ALSSMs

We note that, if two ALSSMs share common model parameters  $A$  and  $c$ , and are supplemented by a common window  $w$ , then they also share their feature vector space  $\mathcal{F}$ , cf. Section 4.3. As an immediate consequence of the Euclidean feature vector space introduced in Section 4.3.3, the Euclidian distance between the output sequences of two such systems is efficiently computed using transformed feature vectors.

Given are two ALSSMs, both of order  $N \in \mathbb{N}$  and with common fixed parameters  $A \in \mathbb{R}^{N \times N}$  and  $c \in \mathbb{R}^{1 \times N}$ , individual initial states  $x \in \mathbb{R}^{N \times 1}$  and  $x' \in \mathbb{R}^{N \times 1}$ , and output sequences  $y$  with elements  $y_k = cA^k x$  and  $y'$  with elements  $y'_k = cA^k x'$  and a window  $w$  as in Section 5.1. Then

the localized Euclidean distance writes as

$$\begin{aligned}
\|y - y'\|_w^2 &= \sum_{i=a}^b w_i \|cA^k x - cA^k x'\|^2 \\
&= \sum_{i=a}^b w_i (x - x')^\top (A^k)^\top c^\top c A^k (x - x') \\
&= (x - x')^\top \underbrace{\left( \sum_{i=a}^b w_i (A^k)^\top c^\top c A^k \right)}_{W = V^\top V} (x - x') \\
&= \|Vx - Vx'\|^2 \tag{5.6} \\
&= \|z - z'\|^2, \tag{5.7}
\end{aligned}$$

where in step (5.6) the Cholesky decomposition  $W = V^\top V$  is used ( $W$  is a positive semi-definite matrix, cf. also Section 4.3.3). Note that  $W$  (and also  $V$ ) does not depend on the state vectors  $x$  or  $x'$  and, thus, the decomposition is preferably precalculated.

## 5.4 Lossless Multi-Step Model Fitting

As applied later in Chapter 7 and in [50], it is often computationally advantageous to split up fitting tasks involving complex models into multiple consecutive steps. In a first step the input signal is fit to an intermediate model, while the later step(s) fit the advanced model(s) onto the feature vectors of the preceding step(s).

In general, such a multi-step fitting comes with the risk of error accumulation. However, we can avoid such an error accumulation for least square error fits, if the intermediate fitting steps respect the following conditions: first, all fitting steps with exception of the last one are projections to the model's feature vector space, i.e., their step's input signals are fit unconstrained. Second, each step's feature vector space is a superset of the next step's feature vector space. It then follows that splitting up a task does not modify the final result.

**Proposition 2.** *Let  $V$  be a Hilbert Space,  $U$  a subspace of  $V$ , and  $S$  a subset of  $U$ , i.e.,  $S \subset U \subset V$ . Then, for any given  $v \in V$*

$$\operatorname{argmin}_{s \in S} \|v - s\|^2 = \operatorname{argmin}_{s \in S} \|u - s\|^2 \tag{5.8}$$



with  $u \in U$  the projection of  $v$  to  $U$ .

*Proof.* Let  $v \in V$ ,  $s \in S$  with  $s = \operatorname{argmin}_{s \in S} \|v - s\|^2$ , and  $s' \in S$  with  $s' \neq s$ , i.e.,

$$\|v - s'\|^2 \geq \|v - s\|^2. \quad (5.9)$$

Then, with  $u \in U$  the projection of  $v$  to  $U$ , it follows that

$$\begin{aligned} \|v - u + u - s'\|^2 &\geq \|v - u + u - s\|^2 \\ \|v - u\|^2 + \|u - s'\|^2 &\geq \|v - u\|^2 + \|u - s\|^2 \\ \|u - s'\|^2 &\geq \|u - s\|^2 \end{aligned} \quad (5.10)$$

where in (5.10) the *Pythagoras' Theorem* is applied. □

**Example 1** (Fit of multi-channel signal to common polynomial). Let  $y_k^{(m)} \in \mathbb{R}$  be the  $k$ th sample of the  $m$ th channel of an  $M$ -channel discrete-time signal (see Fig. 5.1). We are interested in finding a single discrete-time polynomial of order  $Q$  that fits over the time interval  $k \in [a, b]$  the given samples of the  $M$  channels best (in terms of least square error). Therefore, we look for

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{m=1}^M \sum_{k=a}^b \left[ \alpha^\top x^q - y_k^{(m)} \right]^2 \quad (5.11)$$

with polynomial coefficients  $\alpha \in \mathbb{R}^Q$ , polynomial exponents  $q \in \mathbb{N}_0^Q$ . Note that we here already use the vector exponent notation where  $x^q = [x^{q_1}, x^{q_2}, \dots, x^{q_Q}]^\top$  with  $q_i$  the  $i$ th element of  $q$ , as introduced later on in (6.1) in Chapter 6.

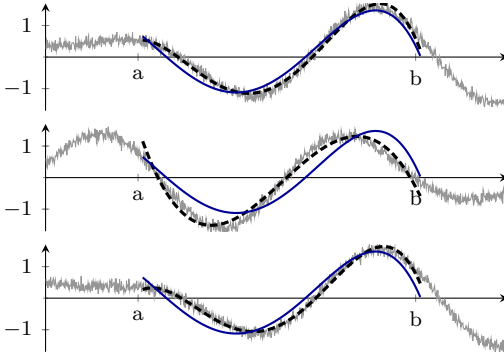
To simplify this minimization, we decompose the task into a first step, a channel-wise fit to a polynomial model,

$$\tilde{\alpha}_m = \operatorname{argmin}_{\alpha} \sum_{k=a}^b \left[ \alpha^\top x^q - y_k^{(m)} \right]^2, \quad (5.12)$$

and into a second step, constraining all channel-wise polynomials from the first step to be the same,

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{m=1}^M \sum_{k=a}^b \left[ \alpha^\top x^q - \tilde{\alpha}_m^\top x^q \right]^2. \quad (5.13)$$

We note that the set of all multi-channel polynomials from the first step with coefficients  $(\tilde{\alpha}_1, \dots, \tilde{\alpha}_M)$  form a *subspace* of the space  $\mathbb{R}^{N \times M}$  spanned by all samples, whereas polynomials from the second step with common coefficients  $(\hat{\alpha}, \dots, \hat{\alpha})$  form a *subset* of that subspace (in this case the subset is even a *subspace*). It follows that, according to Proposition 2, the result of (5.13) is equal to the result of (5.11). ◇



**Figure 5.1:** Local signal approximation of a multi-channel signal over the interval  $\{a, \dots, b\}$  according to (5.11). The plot shows multi-channel discrete-time signals  $y_k^{(1)} \dots y_k^{(3)}$  (solid, gray lines); per-channel polynomial approximation with polynomial coefficients  $\tilde{\alpha}_1 \dots \tilde{\alpha}_3$  (dashed lines); polynomial approximation in common for all channels with polynomial coefficients  $\hat{\alpha}$  (solid, dark blue lines).

## 5.5 ALSSM Signal Clustering in a Transformed Feature Space

Blind identification and clustering of repetitive intervals in a signal is a general problem in signal processing. In particular, when processing biological signals such as ECGs (electrocardiograms), EMGs (electromyograms), ENGs (electroneurograms), etc., repetitive pulses of distinct shapes are observed, but their shapes are not known a priori. This is due to the natural variations of biological systems and the variability in sensor locations. Nevertheless, we are interested in blindly identifying these repetitive signal parts, i.e., without knowing their shapes a priori.

Let  $y \in \mathbb{R}^K$  be a single-channel signal of  $K$  samples and  $\mathcal{F}$  be a feature vector space defined by an ALSSM with parameters  $A \in \mathbb{R}^{N \times N}$  and  $c \in \mathbb{R}^{1 \times N}$  and the window  $w \in \mathbb{R}^K$  as proposed in Section 4.3.2. For efficient comparison of signal intervals, we first extract the feature vectors  $\hat{x}_k \in \mathbb{R}^N$  of  $y$  for any index  $k$  (cf. Section 4.4) and map them to the Euclidean feature vectors  $\hat{z}_k \in \mathbb{R}^N$  (cf. Section 4.3.3). Then, we apply a conventional clustering method onto the transformed feature vectors, such as k-means or the like.

Note that the generalization of this approach to multi-channel signals is straightforward (e.g., for  $M$  channels by stacking the single-channel feature vectors  $x_k \in \mathbb{R}^N$  to a multi-channel feature vector  $x_k \in \mathbb{R}^{NM}$ ).

**Example 2** (Clustering of noisy pulse signal). We observe a two-channel signal  $y$  containing multiple spikes (e.g., neuronal spikes) of different origins and, thus, of different shapes (see Fig. 5.2). The spikes are approximately of duration  $L = 20$  samples. To identify and cluster these pulses, we first project the signal for any index  $k$  to a local feature space spanned by the ALSSM

$$A = 0.9 \begin{bmatrix} \cos(2\pi/L) & -\sin(2\pi/L) \\ \sin(2\pi/L) & \cos(2\pi/L) \end{bmatrix}, \quad c = [1, 0], \quad x_0 = [0, \lambda]^\top,$$

which is a fast decaying cosine pulse of amplitude  $\lambda \in \mathbb{R}$  (to be estimated), and with a degenerated window ALSSM

$$\check{A} = [\gamma], \quad \check{c} = [1], \quad \check{x}_0 = [1],$$

with a decay factor  $\gamma = 0.98$ , which is a simple exponentially decaying window. Then we transform the resulting features to the Euclidean feature space and sort them into 4 clusters with the method of k-means. Figure 5.2 shows a synthetic signal sequence with two channels and three different types of pulse patterns and the results after applying our approach. Figure 5.3 shows the cluster assignment over all sample indices.

◇

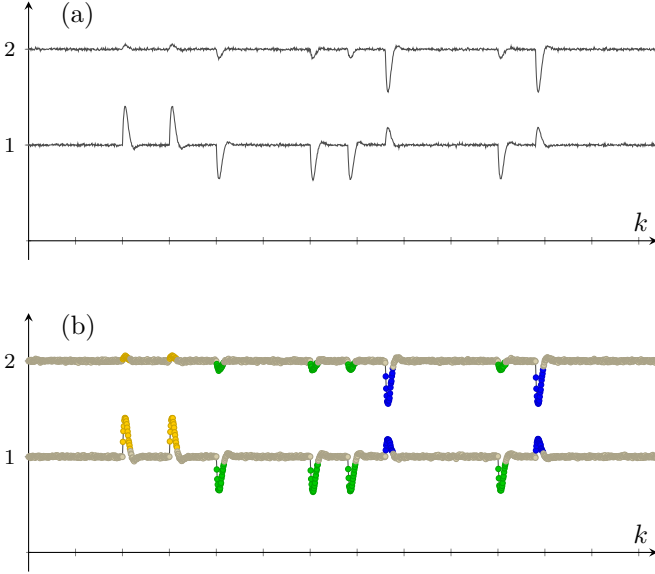
**Example 3** (Clustering pulses of different duration). In this example, we identify and sort pulses of similar shapes but variable duration into 3 clusters. Therefore, we first transform the given 2-channel signal into a feature space spanned by a second order polynomial with a right-sided, exponentially decaying window, cropped at a finite length of 50 samples; the ALSSM parameters are

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \quad c = [1, 0, 0], \quad x_0 = [0, a_1, a_2]^\top,$$

which lead to a discrete-time second order polynomial  $P(x) = a_0 + a_1x + a_2x^2$  with  $a_0$  fixed to 0 and with the two unknown coefficients  $a_1 \in \mathbb{R}$  and  $a_2 \in \mathbb{R}$  (to be estimated). Note fixing  $a_0 = 0$  avoids a jump at the model's signal onset (which is what we expect from our pulses). The parameters for the associated window ALSSM are

$$\check{A} = [\gamma], \quad \check{c} = [1], \quad \check{x}_0 = [1],$$

with a decay factor  $\gamma = 0.98$ , and within a finite support interval with borders  $a = 0$  and  $b = 50$ . To localize the onset of the pulses, we use the LCR (4.82),



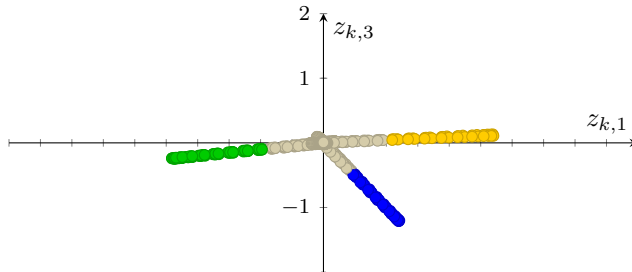
**Figure 5.2:** Blindly identified pulses in a two-channel signal of synthetic neuronal spikes as in Example 2. The upper plot (a) shows the two-channel signal with pulses of damped cosine pulses but of three different amplitudes per channel. The lower plot (b) shows the assignment of each index  $k$  to one out of four clusters according to its local surrounding: clusters 2 to 4 (colored dots in yellow/green/blue) mark the three different pulses, cluster 1 (grey) contains all remaining samples with low signal energy. The feature vectors and their clustering is shown in details in Fig. 5.3.

which here is

$$\text{LCR}_k = -\frac{1}{2} \log \frac{\min_{x \in \Omega} J_k(x)}{J_k(\mathbf{0})}, \quad (5.14)$$

with  $\Omega = \{[0, a_1, a_2]^\top \mid a_1 \in \mathbb{R}, a_2 \in \mathbb{R}\}$  and with the cost function

$$\begin{aligned} J_k(x) &= \sum_{i=a}^b (\check{c}\check{A}^{i-k}\check{x}_0)(cA^{i-k}x - y_i)^2 \\ &= x^\top W_k x - 2x^\top \xi_k + \kappa_k. \end{aligned}$$



**Figure 5.3:** Feature vectors  $z_{k,1}$  and  $z_{k,3}$  of the pulse signal as in Example 2 and Fig. 5.2 assigned to 4 clusters. The colors represent the cluster assignments after applying a k-means algorithm, whereof the gray cluster contains all samples of low energy, i.e., basically the baseline. (Note that  $z_{k,2}$  and  $z_{k,4}$  are all zero as we are only looking for cosine shapes)

according to (4.17) and (4.39). To express the constrained minimization in (5.14) over  $x \in \Omega$  we apply the linear constraint

$$x = Hv \quad \text{with} \quad H = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad v \in \mathbb{R}^2 \quad (5.15)$$

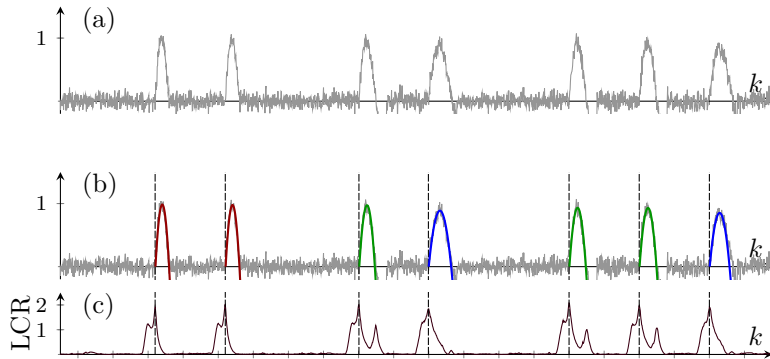
and use the constrained minimization (4.12)

$$\hat{v} = (H^T W_k H)^{-1} H^T \xi_k .$$

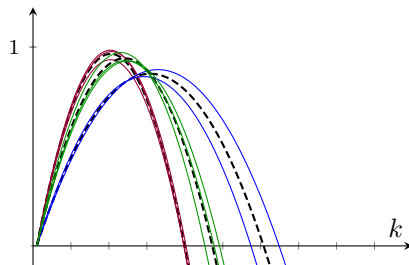
Then, the pulse onsets are most likely located at the maxima of the LCR. To finally sort the detected pulses into different clusters, we transform the feature vectors  $\hat{x} = H\hat{v}$  into an Euclidean feature space (cf. Section 5.5) and apply some common clustering algorithm; for this example we used the k-means algorithm.

Figure 5.4 shows the result of the pulse detection and clustering on a noisy two-channel signal and Fig. 5.5 displays the ALSSM trajectories according to the clustered feature vectors.

◇



**Figure 5.4:** Pulse identification and clustering for a signal as in Example 3. Plot (a) shows a synthetic single-channel signal with pulses of common shapes but of 3 different durations. Plot (b) depicts the identified pulses with their cluster assignments (red/green/blue lines) according to three different pulse durations and Plot (c) shows the detection of the pulse onset, localized at the local maxima of the log-cost ratio (LCR).



**Figure 5.5:** Identified and clustered pulses (red/green/blue) from Fig. 5.4. The cluster centers are superimposed (dashed lines).

# Chapter 6

## Polynomial Cost Functions

### 6.1 Introduction

We have used ALSSMs to locally approximate discrete-time signals in Chapter 4 and to approximate the output of another ALSSM in Chapter 5, both by the means of squared error minimization.

In this chapter, we approximate and substitute discrete-time signals by continuous-time polynomials and use those polynomials to solve complex optimization problems. Substituting discrete-time signals by localized polynomials has shown to be practical, not only in terms of computational efficiency, but also when sub-sample resolution is required or the discrete-time samples are non-equidistant. Polynomials are also useful when multiple variables are involved, leading to multi-variate polynomials. Dealing with polynomial signal approximations in quadratic cost terms is very convenient, since quadratic costs are of a polynomial shape themselves.

Substituting a signal by localized polynomials is not a new idea. Already in 1946, Schoenberger proposed to interpolate (in case of fixed samples) or approximate (in case of samples afflicted with some uncertainty) signals by piece-wise polynomials, i.e., splines [42]. Schoenberger also introduced the now well-known B-splines, a beneficial representation of splines, which is not only nice to handle numerically, but also efficient to evaluate [12]. Nevertheless, the idea of splines later lost its attrac-

tion in the field of signal processing when Shannon's sampling theory took off in the 1950s [44], but remained in focus in fields of mechanics and computer aided design, with particular interest in the surface (2D) modeling [13]. These days, the benefits of splines has been rediscovered for signal processing, as well as for image processing, as emphasized by Unser in 1999 [44].

In the past, polynomials were mostly applied to equidistant (usual case) or non-equidistant (rare case) sampled signals with predefined polynomial borders, i.e., fixed spline knots, and with a rectangular window, weighting each sample equally, when doing the approximation. In our work, we additionally allow non-rectangular, e.g., decaying, windows when approximating a signal. We use such approximations to state application-specific cost functions, commonly by means of least square errors. In such cost functions, we sometimes also allow linear transforms (shifting and dilation) to the time axis of the polynomials (and their windows), to gain optimal shape match. The implementation of applications then often leads to demanding cost functions.

Due to the following interesting properties, polynomials have shown to be practical signal models to work with:

- the *sum* of two polynomials is a polynomial,
- the *product* of two polynomials is a polynomial,
- scaling a polynomial in time (*dilation*) is still a polynomial (of the same order),
- the *derivative* and the *integral* of a polynomial is a polynomial,
- the *composition of two polynomials*, which is obtained by substituting a variable of the first polynomial by the second polynomial, is a polynomial,
- the *minimization of polynomials and rational functions* is a standard problem and supported by all numerical libraries,
- and finally, as mentioned before, a quadratic cost term itself is already of a polynomial form.

Thanks to these properties, least square cost-terms involving polynomial signal approximations are often also of a convenient polynomial form.

To facilitate algebraic transformations and simplifications of such polynomial cost terms, we provide in this chapter our own calculus for polynomials.



A simplistic example illustrating the idea, fitting a multi-channel signal by a single polynomial, was already given in Example 1 in Section 5.4. Multiple examples demonstrating the calculus will be given later in Section 6.3. Finally, non-trivial examples in the field of cardiology, which successfully apply cost functions of polynomial forms, are given in Section 7.3 and in [50].

## 6.2 A Calculus for Polynomials

The use of polynomials in quadratic (or higher order) cost terms soon leads to complex cost functions. Algebraic transformations are needed not only to simplify the terms, but also facilitate precomputations and optimizations using standard libraries. These transformations are often cumbersome, in particular when multivariate terms are involved. Therefore, we here propose a vector-based notation for polynomials, along with a set of manipulation rules and conclude with several examples.

### 6.2.1 Vector Exponent Notation

There are many ways to represent univariate or multivariate polynomials. Probably most common is the writing of a polynomial of order  $Q$  as a sum  $a_0 + a_1x + a_2x^2 + \dots + a_Qx^Q$  or, using the summation notation, as  $\sum_{i=0}^Q a_i x^i$ , where we have the scalar coefficients  $a_i$  and the scalar indeterminate  $x$ . But there are alternative representation forms using coefficient vectors as, for example, in [17], which are often more convenient, once one gets used to them. We propose and explore in the following such an alternative representation form, which uses scalars to the power of vectors, denoted as *vector exponents*. For multivariate polynomials, the Kronecker product is used to flatten multidimensional matrices, cf. Section (6.3.4) and [28].

Let  $x$  be a scalar in  $\mathbb{R}$  and  $q = [q_1, \dots, q_Q]^T$  a vector in  $\mathbb{N}_0^Q$ ,  $Q \in \mathbb{N}$ . Then we define the function  $(x, q) \mapsto x^q$  as the vector of powers of  $x$  by each element of  $q$ , i.e.,

$$x^q \triangleq [x^{q_1}, \dots, x^{q_Q}]^T. \quad (6.1)$$

Note that the elements of  $q$  are neither ordered nor unique. These two properties are essential for an efficient calculus.

## 6.2.2 Univariate Polynomials

The use of vector exponents leads to the following properties for univariate polynomials:

1. The inner product between the coefficient vector  $\alpha \in \mathbb{R}^Q$  and  $x^q$  from (6.1) is an univariate polynomial in  $x$  of order  $\max(q)$ ,

$$\alpha^\top x^q \in \mathbb{R} . \quad (6.2)$$

**Example 1** (A polynomial as an inner product).  $a_0x^0 + a_1x^1 + a_2x^3 = \alpha^\top x^q$  is true for  $\alpha^\top = [a_0, a_1, a_2]$  and  $q = [0, 1, 3]^\top$ .  $\diamond$

2. The sum of such two polynomials with  $\alpha \in \mathbb{R}^Q$  and  $\beta \in \mathbb{R}^Q$  is

$$\alpha^\top x^q + \beta^\top x^q = (\alpha + \beta)^\top x^q \in \mathbb{R} . \quad (6.3)$$

**Example 2** (Sum of Polynomials).  $(a_0x^1 + a_1x^3) + (b_0x^1 + b_1x^3) = (\alpha + \beta)^\top x^q$  is true for  $\alpha = [a_0, a_1]^\top$ ,  $\beta = [b_0, b_1]^\top$  and  $q = [1, 3]^\top$ .  $\diamond$

3. The sum of such two polynomials with  $\alpha \in \mathbb{R}^Q$  and  $\beta \in \mathbb{R}^R$ , with exponent vectors  $q = [q_1, \dots, q_Q]$  and  $r = [r_1, \dots, r_R]$ , and  $Q, R \in \mathbb{N}_0$  is

$$\alpha^\top x^q + \beta^\top x^r = \tilde{\alpha}^\top x^{\tilde{q}} \in \mathbb{R} , \quad (6.4)$$

with

$$\tilde{\alpha} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \in \mathbb{R}^{(Q+R)} \quad (6.5)$$

$$\tilde{q} = \begin{bmatrix} q \\ r \end{bmatrix} \in \mathbb{R}^{(Q+R)} . \quad (6.6)$$

4. The product of two polynomials with common exponent vector  $q$  and coefficients  $\alpha \in \mathbb{R}^Q$  and  $\beta \in \mathbb{R}^Q$  is

$$\begin{aligned} (\alpha^\top x^q)(\beta^\top x^q) &= (\alpha \otimes \beta)^\top (x^q \otimes x^q) \\ &= \tilde{\alpha}^\top x^{\tilde{q}} \end{aligned} \quad (6.7)$$

with

$$\tilde{\alpha} = \alpha \otimes \beta \in \mathbb{R}^{Q^2} \quad (6.8)$$

$$\tilde{q} = Mq \in \mathbb{N}_0^{Q^2} , \quad (6.9)$$

and with

$$M = (I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q) \quad (6.10)$$

where  $I_n$  is the identity matrix and  $\mathbf{1}_n = [1, \dots, 1]^\top$ , both of size  $n$ . The full derivation is given in Appendix (B.1) on page 117.

5. The square of a polynomial is

$$(\alpha^\top x^q)^2 = \tilde{\alpha}^\top x^{\tilde{q}} \quad (6.11)$$

with

$$\tilde{\alpha} = \alpha \otimes \alpha \in \mathbb{R}^{Q^2}, \quad (6.12)$$

$$\tilde{q} = Mq \in \mathbb{N}_0^{Q^2}, \quad (6.13)$$

and with  $M$  as in (6.10). Obviously, this is a special case of polynomial multiplication (6.7).

6. The product of two polynomials with distinct exponent vectors  $q \in \mathbb{N}_0^Q$  and  $r \in \mathbb{N}_0^R$ , and coefficients  $\alpha \in \mathbb{R}^Q$  and  $\beta \in \mathbb{R}^R$ , is

$$(\alpha^\top x^q)(\beta^\top x^r) = \tilde{\alpha}^\top x^{\tilde{q}} \quad (6.14)$$

with

$$\tilde{\alpha} = \alpha \otimes \beta \in \mathbb{R}^{QR}, \quad (6.15)$$

$$\tilde{q} = (I_Q \otimes \mathbf{1}_R)q + (\mathbf{1}_Q \otimes I_R)r \in \mathbb{N}_0^{QR}. \quad (6.16)$$

The derivation is in analogy to the derivation given in Appendix B.1 on page 117.

7. The integral of a polynomial with  $\alpha \in \mathbb{R}^Q$  is

$$\begin{aligned} \int (\alpha^\top x^q) dx &= \alpha^\top \int x^q dx \\ &= \tilde{\alpha}^\top x^{\tilde{q}}. \end{aligned} \quad (6.17)$$

with

$$\tilde{\alpha} = \Lambda \alpha \in \mathbb{R}^Q, \quad (6.18)$$

$$\tilde{q} = q + \mathbf{1}_Q \in \mathbb{N}_0^Q, \quad (6.19)$$

and with

$$\Lambda = \text{diag}(\delta) \in \mathbb{R}^{Q \times Q}, \quad (6.20)$$

where  $\delta \in \mathbb{R}^Q$  and

$$\delta_i = \frac{1}{q_i + 1}, \quad i \in \{1, \dots, Q\} \quad (6.21)$$

with  $\delta_i$  denoting the  $i$ th element of vector  $\delta$  and  $q_i$  the  $i$ th element of vector  $q$ .

8. Definite integral of polynomial with  $\alpha \in \mathbb{R}^Q$ :

$$\begin{aligned} \int_a^b (\alpha^\top x^q) dx &= (\Lambda \alpha)^\top (b^{\tilde{q}} - a^{\tilde{q}}) \\ &= \tilde{\alpha}^\top x^0 \end{aligned} \quad (6.22)$$

with

$$\tilde{\alpha} = (\Lambda \alpha)^\top (b^{\tilde{q}} - a^{\tilde{q}}) \in \mathbb{R}, \quad (6.23)$$

and  $\tilde{q}$  as in (6.19) and with  $\Lambda$  as in (6.20).

9. The derivative of a polynomial with  $\alpha \in \mathbb{R}^Q$  is

$$\begin{aligned} \frac{d}{dx} (\alpha^\top x^q) dx &= \alpha^\top \frac{d}{dx} x^q \\ &= \tilde{\alpha}^\top x^{\tilde{q}} \end{aligned} \quad (6.24)$$

with

$$\tilde{\alpha} = \Lambda \alpha \in \mathbb{R}^Q, \quad (6.25)$$

$$\tilde{q} = \max(q - \mathbf{1}_Q, \mathbf{0}_Q) \in \mathbb{N}_0^Q, \quad (6.26)$$

where  $\max(v_1, v_2)$  denotes the element-wise maxima of corresponding elements in vectors  $v_1$  and  $v_2$ , and with

$$\Lambda = \text{diag}(q). \quad (6.27)$$

10. Shifting a polynomial by a constant value  $\gamma \in \mathbb{R}$  leads to

$$\alpha^\top (x + \gamma)^q = \tilde{\alpha}^\top x^{\tilde{q}} \quad (6.28)$$

with

$$\tilde{\alpha} = \Lambda \alpha \in \mathbb{R}^{\tilde{Q}}, \quad (6.29)$$

$$\tilde{q} = [0, \dots, \max(q)]^T \in \mathbb{N}_0^{\tilde{Q}}, \quad (6.30)$$

$$\tilde{Q} = \max(q) + 1, \quad (6.31)$$

and with

$$\Lambda \in \mathbb{R}^{\tilde{Q} \times Q},$$

where, for all  $i \in [1, \dots, Q]$  and  $j \in [1, \dots, \tilde{Q}]$ ,

$$\Lambda_{\{j,i\}} = \begin{cases} \frac{q_i!}{\tilde{q}_j!(q_i - \tilde{q}_j)!} \gamma^{(q_i - \tilde{q}_j)} & \text{for } \tilde{q}_j \leq q_i, \\ 0 & \text{otherwise.} \end{cases} \quad (6.32)$$

Note that the scalar  $\Lambda_{\{j,i\}}$  refers to the element in the  $j$ th row and the  $i$ th column of  $\Lambda$  and is also denoted as a coefficient of Bernoulli polynomials.

11. The dilation of a polynomial by constant value  $\eta \in \mathbb{R}$  is

$$\alpha^T (\eta x)^q = \tilde{\alpha}^T x^q \quad (6.33)$$

with

$$\tilde{\alpha} = \eta^q \circ \alpha = \Lambda \alpha \in \mathbb{R}^Q, \quad (6.34)$$

and with

$$\Lambda = \text{diag}(\eta^q). \quad (6.35)$$

### 6.2.3 Multivariate Polynomials

To extend our calculus to multivariate polynomials, we for now restrict to bivariate examples; note that the generalization to multivariate polynomials is most often straightforward, e.g., by applying the calculus rules repeatedly.

Bivariate polynomials in  $x$  and  $y$ , both scalars in  $\mathbb{R}$ , write as

$$\alpha^T (x^r \otimes y^q) \in \mathbb{R} \quad (6.36)$$

with  $r \in \mathbb{N}_0^R$  and  $q \in \mathbb{N}_0^Q$ ,  $R$  and  $Q \in \mathbb{N}$ , and the coefficient vector  $\alpha \in \mathbb{R}^{RQ}$ . Such polynomials have the following properties:

1. The sum of two univariate polynomials is

$$\alpha^\top x^q + \beta^\top y^r = (\tilde{\alpha} + \tilde{\beta})^\top (x^{\tilde{q}} \otimes y^{\tilde{r}}) \quad (6.37)$$

with

$$\tilde{\alpha} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathbf{0}_R \end{bmatrix}, \quad \tilde{q} = \begin{bmatrix} 0 \\ q \end{bmatrix}, \quad (6.38)$$

$$\tilde{\beta} = \begin{bmatrix} 1 \\ \mathbf{0}_Q \end{bmatrix} \otimes \begin{bmatrix} 0 \\ \beta \end{bmatrix}, \quad \tilde{r} = \begin{bmatrix} 0 \\ r \end{bmatrix}, \quad (6.39)$$

with  $\mathbf{0}_n = [0, \dots, 0]^\top$  of size  $n$ . The full derivation to this is given in Appendix B.4 on page 118.

2. The product of two univariate polynomials is

$$(\alpha^\top x^q)(\beta^\top y^r) = (\alpha \otimes \beta)^\top (x^q \otimes y^r) \quad (6.40)$$

with full derivation given in Appendix B.5 on page 119.

3. The product of an univariate and a bivariate polynomial is

$$(\alpha^\top x^q)(\beta^\top (y_1^{r_1} \otimes y_2^{r_2})) = (\alpha \otimes \beta)^\top (x^q \otimes y_1^{r_1} \otimes y_2^{r_2}), \quad (6.41)$$

using a derivation in analogy to the derivation given in Appendix B.5 on page 119.

4. The square of a bivariate polynomial with coefficients decomposed as a Kronecker product, i.e., of the form  $\alpha \otimes \beta$  with  $\alpha \in \mathbb{R}^Q$  and  $\beta \in \mathbb{R}^R$ , is

$$\begin{aligned} [(\alpha \otimes \beta)^\top (x^q \otimes y^r)]^2 &= [(\alpha \otimes \alpha) \otimes (\beta \otimes \beta)]^\top [(x^q \otimes x^q) \otimes (y^r \otimes y^r)] \\ &= ((\alpha \otimes \alpha) \otimes (\beta \otimes \beta))^\top (x^{Mq} \otimes y^{M'r}) \end{aligned} \quad (6.42)$$

with  $M$  and  $M'$  as in (6.10). The derivation of the first line is in analogy to the derivation in Appendix B.1 on page 117, and for the last line (6.65) is used.

5. The square of a multivariate polynomials (with non-factorized coefficients) is

$$\left[ \alpha^\top (x^q \otimes y^r) \right]^2 = \tilde{\alpha}^\top (x^{\tilde{q}} \otimes y^{\tilde{r}}) \quad (6.43)$$

with

$$\tilde{\alpha} = R_{Q;R}(\alpha \otimes \alpha) \in \mathbb{R}^{Q^2 R^2} , \quad (6.44)$$

$$\tilde{q} = Mq \in \mathbb{R}^{Q^2} , \quad (6.45)$$

$$\tilde{r} = M'r \in \mathbb{R}^{R^2} , \quad (6.46)$$

and  $R_{Q;R}$  as in (6.103), and  $M$  and  $M'$  as in (6.10). The full derivation is given in Appendix B.2 on page 117.

6. Variable duplicates merge as

$$\alpha^\top(z^s \otimes x^q \otimes x^r \otimes y^q) = \alpha^\top(z^s \otimes x^{\tilde{q}} \otimes y^q) \quad (6.47)$$

with

$$\tilde{q} = (\mathbf{1}_R \otimes I_Q)q + (I_R \otimes \mathbf{1}_Q)r \quad (6.48)$$

according to (6.65).

7. A polynomial with variable shift  $y \in \mathbb{R}$  is

$$\alpha^\top(x + y)^q = \tilde{\alpha}^\top(x^{\tilde{q}} \otimes y^{\tilde{q}}) \quad (6.49)$$

with

$$\tilde{\alpha} = \Lambda\alpha \in \mathbb{R}^{\tilde{Q}^2} , \quad (6.50)$$

$$\tilde{q} = [0, \dots, \max(q)]^\top \in \mathbb{N}_0^{\tilde{Q}} , \quad (6.51)$$

and with  $\tilde{Q} = \max(q) + 1$  and

$$\Lambda = \begin{bmatrix} G_1(q, \tilde{q}) \\ \vdots \\ G_{\tilde{Q}}(q, \tilde{q}) \end{bmatrix} \in \mathbb{R}^{\tilde{Q}^2 \times Q} , \quad (6.52)$$

where, for all  $i \in [1, \dots, Q]$  and  $j \in [1, \dots, \tilde{Q}]$ , and  $n \in \mathbb{N}_0$ ,

$$[G_n(r, s)]_{\{j, i\}} = \begin{cases} \frac{r_i!}{s_j!} \frac{1}{(r_i - s_j)!} & \text{for } r_i - s_j = n \text{ ,} \\ 0 & \text{otherwise.} \end{cases} \quad (6.53)$$

**Example 3** (Polynomial with variable shift).

$$\begin{aligned}
& a_0 + a_1(x + y)^2 \\
&= [a_0]^\top (x + y) \begin{bmatrix} 0 \\ 2 \end{bmatrix} \\
&= \underbrace{(\Lambda \begin{bmatrix} a_0 \\ a_1 \end{bmatrix})^\top}_{\tilde{\alpha} \in \mathbb{R}^9} (x \tilde{q} \otimes y \tilde{q}) \text{ with } \tilde{q} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \text{ and,} \\
&\quad \text{with } \Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \end{bmatrix}^\top, \text{ cf. (6.52).} \\
&= [a_0, 0, a_1, 0, 2a_1, 0, a_1, 0, 0] (x \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \otimes y \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}) \\
&= [a_0, 0, a_1, 0, 2a_1, 0, a_1, 0, 0] \\
&\quad \times [x^0 y^0, x^0 y^1, x^0 y^2, x^1 y^0, x^1 y^1, x^1 y^2, x^2 y^0, x^2 y^1, x^2 y^2]^\top \\
&= a_0 + a_1 y^2 + 2a_1 x y + a_1 x^2 = \tilde{\alpha}^\top (x \tilde{q} \otimes y \tilde{q})
\end{aligned}$$

◇

8. A polynomial dilated by a constant  $\eta \in \mathbb{R}$  is

$$\begin{aligned}
\alpha^\top (z^s \otimes (\eta x)^q \otimes y^r) &= \alpha^\top (z^s \otimes (\eta^q \circ x^q) \otimes y^r) \\
&= (\alpha \circ (\mathbf{1}_S \otimes \eta^q \otimes \mathbf{1}_R))^\top (z^s \otimes x^q \otimes y^r) \\
&= \underbrace{(\text{diag}(\mathbf{1}_S \otimes \eta^q \otimes \mathbf{1}_R) \alpha)}_{\Lambda}^\top (z^s \otimes x^q \otimes y^r) \quad (6.54) \\
&= (\Lambda \alpha)^\top (z^s \otimes x^q \otimes y^r) \quad (6.55)
\end{aligned}$$

using (6.116) in the second step.

9. The dilation of a polynomial by an indeterminate  $y \in \mathbb{R}$  is

$$\alpha^\top (y x)^q = \alpha^\top (y^q \circ x^q) = \text{vdiag}(\alpha)^\top (y^q \otimes x^q) \quad (6.56)$$

with

$$\text{vdiag}(\alpha) \triangleq \text{vec}(\text{diag}(\alpha)) . \quad (6.57)$$

10. Substituting variables by constants, e.g.,  $c \in \mathbb{R}$  is given and known, leads to

$$\alpha^\top (x^q \otimes c^s \otimes y^r) = (\tilde{\Lambda} \alpha)^\top (x^q \otimes y^r) \quad (6.58)$$

with

$$\tilde{\Lambda} = (I_Q \otimes (c^s)^\top \otimes I_R) . \quad (6.59)$$

The full derivation for this is given in Appendix B.6 on page 119.



11. The integral of a multivariate polynomial with respect to the scalar variable  $x$  is given by

$$\int \alpha^\top(z^s \otimes x^q \otimes y^r) dx = (\alpha \circ (1^s \otimes \delta \otimes 1^r))^\top(z^s \otimes x^{\tilde{q}} \otimes y^r) \quad (6.60)$$

with  $\delta \in \mathbb{R}^Q$  as in (6.21) and with  $\tilde{q}$  as in (6.19). The full derivation is given in Appendix B.7 on page 119.

In some applications  $\alpha$  is given in a factorized form  $\alpha = \Lambda \alpha'$  with an appropriate matrix  $\Lambda$  and vector  $\alpha'$ . Then we get, using (6.121),

$$\int \alpha^\top(z^s \otimes x^q \otimes y^r) dx = \underbrace{(\text{diag}(1^s \otimes \delta \otimes 1^r) \Lambda \alpha')}_{\tilde{\Lambda}}^\top(z^s \otimes x^{\tilde{q}} \otimes y^r). \quad (6.61)$$

12. The definite integral of a multivariate polynomial with respect to the scalar variable  $x$  is given by

$$\begin{aligned} \int_a^b \alpha^\top(z^s \otimes x^q \otimes y^r) dx &= (\Lambda \alpha)^\top((z^s \otimes b^{\tilde{q}} \otimes y^r) - (z^s \otimes a^{\tilde{q}} \otimes y^r)) \\ &= \underbrace{((I_s \otimes (b^{\tilde{q}} - a^{\tilde{q}})^\top \otimes I_r) \Lambda \alpha)^\top}_{\tilde{\Lambda}}(z^s \otimes y^r) \end{aligned} \quad (6.62)$$

$$= (\tilde{\Lambda} \alpha)^\top(z^s \otimes y^r) \quad (6.63)$$

with  $\tilde{q}$  as in (6.19) and  $\Lambda$  corresponds to  $\tilde{\Lambda}$  in (6.61). In the last step of the derivation, (6.58) is used.

### Some Useful Operators on Vector Exponents

Let  $x^q$ ,  $x^{q'}$ , and  $x^r$  with  $q, q' \in \mathbb{N}_0^Q$  and  $r \in \mathbb{N}_0^R$  be vectors as defined in (6.1). Then the following relation apply:

1. The Hadamard product (element wise product) of two vector exponents is

$$x^q \circ x^{q'} = x^{q+q'}. \quad (6.64)$$

2. The Kronecker product of two vector exponents of any size is

$$x^q \otimes x^r = x^{\tilde{q}} \quad (6.65)$$

with

$$\tilde{q} = (I_Q \otimes \mathbf{1}_R)q + (\mathbf{1}_Q \otimes I_R)r , \quad (6.66)$$

with full derivation given in Appendix B.3 on page 118.

3. The outer vector product of two vector exponents is

$$x^q(x^r)^\top = (x^r)^\top \otimes x^q \quad (6.67)$$

which directly follows from property (6.106) of Kronecker products.

### 6.2.4 Table Summary

Table 6.1 and Table 6.2 provide a tabular summary of the calculus derived in Section 6.2.2 and Section 6.2.3.

**Table 6.1:** Calculus for Univariate Polynomials ( $x \in \mathbb{R}$ ,  $y \in \mathbb{R}$ ,  $Q \in \mathbb{N}$ ,  $R \in \mathbb{N}$ ,  $\alpha \in \mathbb{R}^Q$ ,  $\beta \in \mathbb{R}^R$ ,  $q \in \mathbb{N}_0^Q$ ,  $r \in \mathbb{N}_0^R$ )

| $f(x) \in \mathbb{R}$               | $f(x) = \tilde{\alpha}^\top x^{\tilde{q}}$   |  |                              | Polyn. order<br>$\hat{=} \max(\tilde{q})$ |
|-------------------------------------|--|--|------------------------------|---|
|                                     | $\tilde{\alpha} \in \mathbb{R}^{\tilde{Q}}$  | $\tilde{q} \in \mathbb{N}_0^{\tilde{Q}}$                     | $\tilde{Q} \in \mathbb{N}_0$ |   |
| $\alpha^\top x^q + \beta^\top x^q$  | $(\alpha + \beta)$   | $q$  | $Q$                          | $\max(q)$                                 |
| $\alpha^\top x^q + \beta^\top x^r$  | $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$                                      | $\begin{bmatrix} q \\ r \end{bmatrix}$                       | $Q + R$                      | $\max(q, r)$                              |
| $(\alpha^\top x^q)^2$               | $\alpha \otimes \alpha$  | $[(I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q)]q$ | $Q^2$                        | $\max(q)^2$                               |
| $(\alpha^\top x^q)(\beta^\top x^q)$ | $\alpha \otimes \beta$   | $[(I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q)]q$ | $Q^2$                        | $\max(q)^2$                               |
| $(\alpha^\top x^q)(\beta^\top x^r)$ | $\alpha \otimes \beta$   | $(I_Q \otimes \mathbf{1}_R)q + (\mathbf{1}_Q \otimes I_R)r$  | $QR$                         | $\max(q) \max(r)$                         |
| $\int (\alpha^\top x^q) dx$         | $\Lambda \alpha,$<br>$\Lambda$ see (6.20)  | $q + \mathbf{1}_Q$   | $Q$                          | $\max(q) + 1$                             |
| $\int_a^b (\alpha^\top x^q) dx$     | $(\Lambda \alpha)^\top (b^{\tilde{q}} - a^{\tilde{q}}),$<br>for $\Lambda$ see (6.20) | 0  | 1                            | $\max(q) + 1$                             |
| $\frac{d}{dx} (\alpha^\top x^q) dx$ | $\text{diag}(q) \alpha$  | $q - \mathbf{1}_Q$   | $Q$                          | $\max(q) - 1$                             |
| $\alpha^\top (x + \gamma)^q$        | $\Lambda \alpha,$ for $\Lambda$ see (6.32)   | $[0, \dots, \max(q)]^\top$                                   | $\max(q) + 1$                | $\max(q)$                                 |
| $\alpha^\top (\eta x)^q$            | $\eta^q \circ \alpha$<br>$= \text{diag}(\eta^q) \alpha$                              | $q$  | $Q$                          | $\max(q)$                                 |

**Table 6.2:** Calculus for Multivariate Polynomials ( $x \in \mathbb{R}$ ,  $y \in \mathbb{R}$ ,  $Q \in \mathbb{N}$ ,  $R \in \mathbb{N}$ ,  $\alpha \in \mathbb{R}^Q$ ,  $\beta \in \mathbb{R}^R$ ,  $q \in \mathbb{N}_0^Q$ ,  $r \in \mathbb{N}_0^R$ )

| $f(x, y, \dots) \in \mathbb{R}$   | $f(x, y, \dots) = \tilde{\alpha}^\top X$  |   |             |
|---|---|---|-------------|
|   | $\tilde{\alpha} \in \mathbb{R}^{\tilde{Q}}$   | $X \in \mathbb{R}^{\tilde{Q}}$  | $\tilde{Q}$ |
| $(\alpha^\top x^q)(\beta^\top y^r)$                                     | $(\alpha \otimes \beta)$  | $(x^q \otimes y^r)$   | $QR$        |
| $(\alpha^\top x^q)(\beta^\top (y_1^{r_1} \otimes y_2^{r_2}))$           | $(\alpha \otimes \beta)$  | $(x^q \otimes y_1^{r_1} \otimes y_2^{r_2})$   | $QR_1R_2$   |
| $\left[ (\alpha \otimes \beta)^\top (x^q \otimes y^r) \right]^2$        | $(\alpha \otimes \alpha) \otimes (\beta \otimes \beta)$   | $(x^{Mq} \otimes y^{M'r})$  | $Q^2R^2$    |
| $\left[ \alpha^\top (x^q \otimes y^r) \right]^2$                        | $R_{Q,R}(\alpha \otimes \alpha),$<br>$R_{Q,R}$ as in (6.102)  | with $M = [(I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q)]$<br>with $M' = [(I_R \otimes \mathbf{1}_R) + (\mathbf{1}_R \otimes I_R)]$ |             |
| $\alpha^\top x^q + \beta^\top y^r$                                      | $(\tilde{\alpha} + \tilde{\beta})^\top (x^{\tilde{q}} \otimes y^{\tilde{r}})$<br>$\tilde{\alpha} = [0 \mid \alpha^\top]^\top \otimes [1 \mid \mathbf{0}_R^\top]^\top$<br>$\tilde{\beta} = [1 \mid \mathbf{0}_Q^\top]^\top \otimes [0 \mid \beta^\top]^\top$<br>$\tilde{q} = [0 \mid q^\top]^\top, \tilde{r} = [0 \mid r^\top]^\top$ | $(x^q \otimes y^r)$   | Q R         |
| $\alpha^\top (z^s \otimes x^q \otimes x^r \otimes y^q)$                 | $\alpha$  | $(z^s \otimes x^{\tilde{q}} \otimes y^q)$<br>with $\tilde{q} = (1^r \otimes q) + (r \otimes 1^q)$   |             |
| $\alpha^\top (x + y)^q$   | $\Lambda \alpha,$ (for $\Lambda$ see (6.52))  | $(x^{\tilde{q}} \otimes y^{\tilde{q}})$<br>with $\tilde{q} = [0, \dots, \tilde{Q} - 1]^\top$  |             |
| $\alpha^\top (yx)^q$  | $\text{vdiag}(\alpha)$  | $(y^q \otimes x^q)$   |             |
| $\int \alpha^\top (\dots \otimes x^q \otimes y^r \otimes \dots) dx$     | $(\alpha \circ (\dots \otimes \delta \otimes 1^r \otimes \dots))$   | $(\dots \otimes x^{\tilde{q}} \otimes y^r \otimes \dots)$   |             |
| $\int_a^b \alpha^\top (\dots \otimes x^q \otimes y^r \otimes \dots) dx$ | $\tilde{\Lambda} \alpha,$ (for $\tilde{\Lambda}$ see (6.62))  | $(\dots \otimes y^q \otimes \dots)$   |             |
| $\alpha^\top (x^q \otimes c^s \otimes y^r)$                             | $\tilde{\Lambda} \tilde{\alpha},$ (cf. (6.58))  | $(x^q \otimes y^r)$   |             |

## 6.3 Applications

### 6.3.1 Squared Error Between Two Polynomials Within a Given Interval

In a first application, we calculate the squared error between two polynomials of any order over a fixed interval. The polynomial coefficients are known but vary depending on observations, e.g., the coefficients are the result from approximating a signal with local polynomials. Thus, the terms should be arranged in such a way, that any term independent of the coefficients can be precalculated off-line.

Let  $\alpha \in \mathbb{R}^Q$  and  $\beta \in \mathbb{R}^Q$  be the coefficients and  $q \in \mathbb{N}_0^Q$  the common exponent vector of two polynomials of order  $\max(q)$ ,  $Q \in \mathbb{N}$ ; we here assume that the two polynomials  $\alpha^\top x^q$  and  $\beta^\top x^q$  are local approximations of observed signals. Thus, we are interested in the squared error  $J(\alpha, \beta) \in \mathbb{R}$  between the two polynomials over the interval  $[a, b]$ . This cost function writes as

$$\begin{aligned}
 J(\alpha, \beta) &= \int_a^b \left[ \alpha^\top x^q - \beta^\top x^q \right]^2 dx \\
 &= \int_a^b \left[ (\alpha - \beta)^\top x^q \right]^2 dx \\
 &= \int_a^b \left( (\alpha - \beta) \otimes (\alpha - \beta) \right)^\top x^{Mq} dx \\
 &= \left[ \left( \Lambda \left( (\alpha - \beta) \otimes (\alpha - \beta) \right) \right)^\top x^{\tilde{q}} \right]_a^b \\
 &= \left( \Lambda \left( (\alpha - \beta) \otimes (\alpha - \beta) \right) \right)^\top (b^{\tilde{q}} - a^{\tilde{q}}) \\
 &= \underbrace{\left( (\alpha \otimes \alpha) - (\beta \otimes \alpha) - (\alpha \otimes \beta) + (\beta \otimes \beta) \right)}_{\text{observation dependant}} \underbrace{\Lambda^\top (b^{\tilde{q}} - a^{\tilde{q}})}_{\text{constant}} \in \mathbb{R}, \quad (6.68)
 \end{aligned}$$

with  $M = [(I_Q \otimes \mathbb{1}_Q) + (\mathbb{1}_Q \otimes I_Q)]$  according to (6.45),  $\Lambda$  according to (6.20), and  $\tilde{q} = Mq + \mathbb{1}$  according to (6.19). Note that in the last line the right part is constant and, hence, fully precomputable.

**Example 4** (Local squared error between two first order polynomials). Let  $\alpha_1 x + \alpha_0$  and  $\beta_1 x + \beta_0$  be two first order polynomials with coefficients  $\alpha = [\alpha_1, \alpha_0]^\top \in \mathbb{R}^2$  and  $\beta = [\beta_1, \beta_0]^\top \in \mathbb{R}^2$ , respectively. Then the squared error

over the given interval  $[a, b]$  is

$$J(\alpha, \beta) = \int_a^b \left[ (\alpha_1 x + \alpha_0) - (\beta_1 x + \beta_0) \right]^2 dx ,$$

or using the vector exponent notation, with  $q = [1, 0]$ ,

$$J(\alpha, \beta) = \int_a^b \left[ \alpha^\top x^q - \beta^\top x^q \right]^2 dx .$$

This is according to (6.68) equivalent to,

$$\begin{aligned} J(\alpha, \beta) &= ((\alpha \otimes \alpha) - (\beta \otimes \alpha) - (\alpha \otimes \beta) + (\beta \otimes \beta))^\top \Lambda^\top (b^{\tilde{q}} - a^{\tilde{q}}) \\ &= \left( ([\alpha_0] \otimes [\alpha_0]) - ([\beta_0] \otimes [\alpha_0]) - ([\alpha_0] \otimes [\beta_0]) + ([\beta_0] \otimes [\beta_0]) \right)^\top \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}^\top (b^{\tilde{q}} - a^{\tilde{q}}) \\ &= \left( \begin{bmatrix} \alpha_1 \alpha_1 \\ \alpha_1 \alpha_0 \\ \alpha_0 \alpha_1 \\ \alpha_0 \alpha_0 \end{bmatrix} - \begin{bmatrix} \beta_1 \alpha_1 \\ \beta_1 \alpha_0 \\ \beta_0 \alpha_1 \\ \beta_0 \alpha_0 \end{bmatrix} - \begin{bmatrix} \alpha_1 \beta_1 \\ \alpha_1 \beta_0 \\ \alpha_0 \beta_1 \\ \alpha_0 \beta_0 \end{bmatrix} + \begin{bmatrix} \beta_1 \beta_1 \\ \beta_1 \beta_0 \\ \beta_0 \beta_1 \\ \beta_0 \beta_0 \end{bmatrix} \right)^\top \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}^\top \left( \begin{bmatrix} b^3 \\ b^2 \\ b^2 \\ b^1 \end{bmatrix} - \begin{bmatrix} a^3 \\ a^2 \\ a^2 \\ a^1 \end{bmatrix} \right) \\ &= \begin{bmatrix} \frac{1}{4} \alpha_1^2 - \frac{1}{2} \alpha_1 \beta_1 + \frac{1}{4} \beta_1^2 \\ \frac{1}{3} \alpha_1 \alpha_0 - \frac{1}{3} \beta_1 \alpha_0 - \frac{1}{3} \alpha_1 \beta_0 + \beta_1 \beta_0 \\ \frac{1}{3} \alpha_1 \alpha_0 - \frac{1}{3} \beta_0 \alpha_1 - \frac{1}{3} \alpha_0 \beta_1 + \frac{1}{3} \beta_1 \beta_0 \\ \frac{1}{2} \alpha_0^2 - \alpha_0 \beta_0 + \frac{1}{2} \beta_0^2 \end{bmatrix}^\top \begin{bmatrix} b^3 - a^3 \\ b^2 - a^2 \\ b^2 - a^2 \\ b^1 - a^1 \end{bmatrix} \end{aligned} \quad (6.69)$$

with

$$\begin{aligned} \tilde{q} &= [(I_2 \otimes \mathbf{1}_2) + (\mathbf{1}_2 \otimes I_2)]q + \mathbf{1}_4 \\ &= \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 2 \\ 1 \end{bmatrix} . \end{aligned}$$

Note that for fixed borders  $a$  and  $b$  the right bracket in (6.69) is also fixed. It remains a single multivariate polynomial in  $\alpha$  and  $\beta$ .

◇

### 6.3.2 Minimum Squared Error Fit of Polynomials by Scaling in its Amplitude and its Time

This application originates from Section 7.3 which addresses the inverse problem in electrocardiography. For that, ECG signals are locally approximated by higher-order (e.g., fifth order) polynomials and then fit to multiple templates; a template here refers to a polynomial approximation of a synthetic ECG sequence generated with a specific set of parameters. To fit a signal to a single template, we minimize the squared error

while we allow the template to scale in its amplitude and time dilation. After this has been done for all templates, we identify the best matching template and thus the best set of parameters.

Note that the ECG signals and the templates used later as in Section 7.3 are of multiple channels. For the sake of simplicity, we here only consider single-channel signals, but the generalization is straightforward.

Let  $\alpha \in \mathbb{R}^Q$  be the polynomial coefficients locally approximating a signal,  $\beta \in \mathbb{R}^Q$  the polynomial coefficients of a template, and  $q \in \mathbb{N}_0^Q$  the vector exponent of both polynomials and of order  $\max(q)$ ,  $Q \in \mathbb{N}$ . We are interested to minimize the squared error  $J(\lambda, \eta) \in \mathbb{R}$  between the two polynomials over the interval  $[a, b]$  by optimizing the template amplitude  $\lambda \in \mathbb{R}$  and its time dilation  $\eta \in \mathbb{R}$ . The comparison considers the fixed interval  $[a, b]$ ,  $a \in \mathbb{R}$ ,  $b \in \mathbb{R}$ ,  $a < b$ . Then the cost function writes as

$$\begin{aligned}
 J(\lambda, \eta) &= \int_a^b \left[ \alpha^\top x^q - \lambda \beta^\top (\eta x)^q \right]^2 dx & (6.70) \\
 &= \left[ \underbrace{(\delta \circ (\alpha \otimes \alpha))^\top}_{c_1} (x^{Mq+1}) \right]_a^b \\
 &\quad - 2\lambda \left[ \underbrace{((1^q \otimes \delta) \circ (\text{vdiag}(\beta) \otimes \alpha))^\top}_{c_2} (\eta^q \otimes x^{Mq+1}) \right]_a^b \\
 &\quad + \lambda^2 \left[ \underbrace{((1^{Mq} \otimes \delta) \circ (R_{Q,Q}(\text{vdiag}(\beta) \otimes \text{vdiag}(\beta))))^\top}_{c_3} \right. \\
 &\quad \quad \left. \cdot (\eta^{Mq} \otimes x^{Mq+1}) \right]_a^b \\
 &= \tilde{\alpha}_1 - 2\lambda \tilde{\alpha}_2^\top \eta^q + \lambda^2 \tilde{\alpha}_3^\top \eta^{Mq} & (6.71)
 \end{aligned}$$

with the coefficient vectors

$$\tilde{\alpha}_1 = (b^{Mq+1} - a^{Mq+1})^\top c_1 \in \mathbb{R}, \quad (6.72)$$

$$\tilde{\alpha}_2 = (I_Q \otimes (b^{Mq+1} - a^{Mq+1})^\top) c_2 \in \mathbb{R}^Q, \quad (6.73)$$

$$\tilde{\alpha}_3 = (I_{Q^2} \otimes (b^{Mq+1} - a^{Mq+1})^\top) c_3 \in \mathbb{R}^{Q^2}, \quad (6.74)$$

and, as in (6.10), with

$$M = [(I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q)]. \quad (6.75)$$

Note that these coefficient vectors  $\tilde{\alpha}_1$ ,  $\tilde{\alpha}_2$ , and  $\tilde{\alpha}_3$  are independent of  $\lambda$ ,  $\eta$  and  $x$  and, thus, precomputable. The full derivation is given in Appendix B.14 on page 120.

Finally, as we have in (6.71) a quadratic form in  $\lambda$ , it is minimized over  $\lambda$  by

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} J(\lambda, \eta) = -\frac{\tilde{\alpha}_2^\top \eta^q}{\tilde{\alpha}_3^\top \eta^{Mq}}. \quad (6.76)$$

Further, substituting  $\lambda$  by  $\hat{\lambda}$  in (6.71) we get

$$\begin{aligned} J(\hat{\lambda}, \eta) &= \left( -\frac{\tilde{\alpha}_2^\top \eta^q}{\tilde{\alpha}_3^\top \eta^{Mq}} \right)^2 \tilde{\alpha}_3^\top \eta^{Mq} - 2 \left( -\frac{\tilde{\alpha}_2^\top \eta^q}{\tilde{\alpha}_3^\top \eta^{Mq}} \right) \tilde{\alpha}_2^\top \eta^q + \tilde{\alpha}_1 \\ &= -\frac{(\tilde{\alpha}_2^\top \eta^q)^2}{\tilde{\alpha}_3^\top \eta^{Mq}} + 2 \frac{(\tilde{\alpha}_2^\top \eta^q)^2}{\tilde{\alpha}_3^\top \eta^{Mq}} + \tilde{\alpha}_1 \\ &= \frac{(\tilde{\alpha}_2^\top \eta^q)^2}{\tilde{\alpha}_3^\top \eta^{Mq}} + \tilde{\alpha}_1, \end{aligned} \quad (6.77)$$

and minimized over  $\eta$ , we get

$$\begin{aligned} \hat{\eta} &= \underset{\eta}{\operatorname{argmin}} J(\hat{\lambda}, \eta) = \underset{\eta}{\operatorname{argmin}} \frac{(\tilde{\alpha}_2^\top \eta^q)^2}{\tilde{\alpha}_3^\top \eta^{Mq}} \\ &= \underset{\eta}{\operatorname{argmin}} \frac{(\tilde{\alpha}_2 \otimes \tilde{\alpha}_2)^\top \eta^{Mq}}{\tilde{\alpha}_3^\top \eta^{Mq}}. \end{aligned} \quad (6.78)$$

Note that (6.78) is a univariate rational function of maximum nominator and denominator order of  $2 \max(q)$  and needs to be minimized with respect to  $\eta \in \mathbb{R}$ , e.g., by applying any standard numerical method.

An immediate application to this result, generalized to multiple channels, is given in Section 7.3, solving the inverse problem of esophageal ECG signals.

### 6.3.3 Offset-Free Minimum Squared Error Fit of Polynomials by Scaling in Amplitude and Time

We extend the application given in Section 6.3.2 and perform an offset-free comparison, i.e., we ignore any constant offset values between the signals and the templates. Thus, we modify (6.70) to

$$J(\lambda, \eta) = \int_a^b \left[ (\alpha^\top x^q - m_\alpha) - \lambda (\beta^\top (\eta x)^q - m_\beta(\eta)) \right]^2 dx \quad (6.79)$$



with the two mean or offset values

$$m_\alpha = \frac{1}{b-a} \int_a^b \alpha^\top x^q dx , \quad (6.80)$$

$$m_\beta(\eta) = \frac{1}{b-a} \int_a^b \beta^\top (\eta x)^q dx , \quad (6.81)$$

which we rewrite using the vector exponent notation as

$$m_\alpha = \alpha^\top \left( \frac{1}{b-a} \text{diag}(\delta) (b^{q+1} - a^{q+1}) \right) , \quad (6.82)$$

$$m_\beta(\eta) = \beta_m^\top \eta^q , \quad (6.83)$$

with

$$\beta_m = \frac{1}{b-a} (I_Q \otimes (b^q - a^q)) (\text{diag}(1^q \otimes \delta) \beta) . \quad (6.84)$$

The full derivation to this is given in Appendix B.9 on page 121.

On the way to minimize  $J(\lambda, \eta)$  in (6.79), we now substitute  $m_\alpha$  and  $m_\beta$  in (6.79), and expand all polynomial terms to a common exponent vector  $\tilde{q}$ , i.e.,

$$J(\lambda, \eta) = \int_a^b \left[ \underbrace{(\tilde{\alpha}^\top x^{\tilde{q}})}_{\alpha^\top x^q} - \underbrace{\tilde{\alpha}_m^\top x^{\tilde{q}}}_{m_\alpha} - \lambda \left( \underbrace{\tilde{\beta}^\top (\eta x)^{\tilde{q}}}_{\beta^\top (\eta x)^q} - \underbrace{\tilde{\beta}_m \eta^{\tilde{q}}}_{\beta_m^\top \eta^q} \underbrace{\alpha_0^\top x^{\tilde{q}}}_1 \right) \right]^2 dx$$

with

$$\begin{aligned} \tilde{q} &= [ 0 \mid q^\top ]^\top & \tilde{\alpha} &= [ 0 \mid \alpha^\top ]^\top & \tilde{\alpha}_m &= [ m_\alpha \mid \mathbf{0}_Q^\top ]^\top , \\ \tilde{\beta} &= [ 0 \mid \beta^\top ]^\top & \tilde{\beta}_m &= [ 0 \mid \beta_m^\top ]^\top & \tilde{\alpha}_0 &= [ 1 \mid \mathbf{0}_Q^\top ]^\top . \end{aligned}$$

Thus, we get

$$\begin{aligned} J(\lambda, \eta) &= \int_a^b \left[ (\alpha - \alpha_m)^\top x^{\tilde{q}} - \lambda (\text{vdiag}(\tilde{\beta})^\top (x^{\tilde{q}} \otimes \eta^{\tilde{q}}) - (\tilde{\beta}_m \otimes \alpha_0)^\top (\eta^{\tilde{q}} \otimes x^{\tilde{q}})) \right]^2 dx \\ &= \int_a^b \left[ \underbrace{(\alpha - \alpha_m)^\top}_{\alpha'} x^{\tilde{q}} - \lambda \underbrace{(\text{vdiag}(\tilde{\beta}) - (\tilde{\beta}_m \otimes \alpha_0))^\top}_{\beta'} (\eta^{\tilde{q}} \otimes x^{\tilde{q}}) \right]^2 dx \\ &= \int_a^b \left[ \alpha'^\top x^{\tilde{q}} - \lambda \beta'^\top (\eta^{\tilde{q}} \otimes x^{\tilde{q}}) \right]^2 dx , \quad (6.85) \end{aligned}$$

which is of the same form as (6.70), or more precisely as (B.11) in Appendix B.8 on page 120, and can be minimized with respect to  $\lambda$  and  $\eta$  accordingly.

### 6.3.4 Delay Estimation Using Local Polynomial Approximations

In this application, we wish to estimate the time delay between two signals of a single source but observed with a per-channel time-dependent delay. This delay will be estimated using local signal approximates and comparing them by means of minimum squared error.

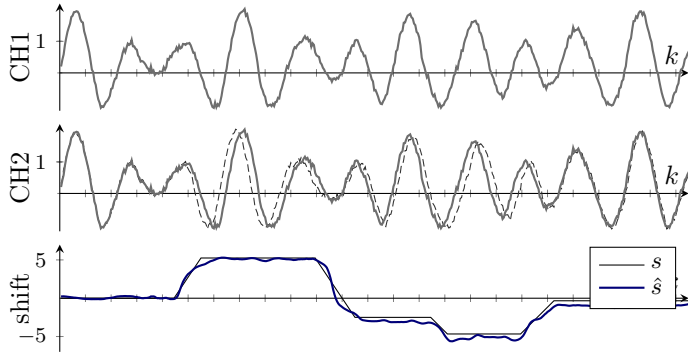
Let  $\alpha \in \mathbb{R}^N$  and  $\beta \in \mathbb{R}^N$  be the polynomial coefficients of the local approximations over the interval  $[a, b]$  of the two observed signals. Then we find the estimate  $\hat{s} \in \mathbb{R}$  of the time-delay between the signals by minimizing the squared error

$$\begin{aligned}
J(s) &= \int_a^b \left[ \alpha^\top \left(x - \frac{1}{2}s\right)^q - \beta^\top \left(x + \frac{1}{2}s\right)^q \right]^2 dx & (6.86) \\
&= \int_a^b \left[ (\Lambda\alpha)^\top \left(x^{\tilde{q}} \otimes \left(-\frac{1}{2}s\right)^{\tilde{q}}\right) - (\Lambda\beta)^\top \left(x^{\tilde{q}} \otimes \left(\frac{1}{2}s\right)^{\tilde{q}}\right) \right]^2 dx \\
&= \int_a^b \left[ (\tilde{\Lambda}\alpha)^\top \left(x^{\tilde{q}} \otimes s^{\tilde{q}}\right) - (\tilde{B}\beta)^\top \left(x^{\tilde{q}} \otimes s^{\tilde{q}}\right) \right]^2 dx \\
&= \int_a^b \left[ (\tilde{\Lambda}\alpha - \tilde{B}\beta)^\top \left(x^{\tilde{q}} \otimes s^{\tilde{q}}\right) \right]^2 dx \\
&= \int_a^b \left( R_{Q;Q} [(\tilde{\Lambda}\alpha - \tilde{B}\beta) \otimes (\tilde{\Lambda}\alpha - \tilde{B}\beta)] \right)^\top \left( x^{\overbrace{M\tilde{q}}^{q'}} \otimes s^{\overbrace{M\tilde{q}}^{q'}} \right) dx \\
&= \left( \tilde{C} [(\tilde{\Lambda}\alpha - \tilde{B}\beta) \otimes (\tilde{\Lambda}\alpha - \tilde{B}\beta)] \right)^\top s^{q'} \\
&= \left( \tilde{C} \left[ (\tilde{\Lambda}\alpha \otimes \tilde{\Lambda}\alpha) - (\tilde{\Lambda}\alpha \otimes \tilde{B}\beta) - (\tilde{B}\beta \otimes \tilde{\Lambda}\alpha) + (\tilde{B}\beta \otimes \tilde{B}\beta) \right] \right)^\top s^{q'} \\
&= \left( \tilde{C} \left[ (\tilde{\Lambda}\alpha \otimes \tilde{\Lambda}\alpha) - \tilde{K}(\tilde{\Lambda}\alpha \otimes \tilde{B}\beta) + (\tilde{B}\beta \otimes \tilde{B}\beta) \right] \right)^\top s^{q'} \\
&= \left( \underbrace{\tilde{C}(\tilde{\Lambda} \otimes \tilde{\Lambda})}_{A} (\alpha \otimes \alpha) - \underbrace{\tilde{C}\tilde{K}(\tilde{\Lambda} \otimes \tilde{B})}_{B} (\alpha \otimes \beta) + \underbrace{\tilde{C}(\tilde{B} \otimes \tilde{B})}_{C} (\beta \otimes \beta) \right)^\top s^{q'} \\
&= (A(\alpha \otimes \alpha) - B(\alpha \otimes \beta) + C(\beta \otimes \beta))^\top s^{q'} , & (6.87)
\end{aligned}$$

with  $\Lambda$  as in (6.52),  $\tilde{q}$  as in (6.51),  $\tilde{\Lambda} = \text{diag}(1^{\tilde{q}} \otimes (-\frac{1}{2})^{\tilde{q}})\Lambda$  according to (6.54),  $\tilde{B} = \text{diag}(1^{\tilde{q}} \otimes (\frac{1}{2})^{\tilde{q}})\Lambda$ ,  $\tilde{C} = ((b^{q'+1} - a^{q'+1})^\top \otimes I_{\tilde{q}})\tilde{C}'$  according to (6.62), and  $\tilde{C}' = (\text{diag}(\delta \otimes 1^{q'+1})R_{q,r})$  according to (6.61), and with  $\tilde{K} = I + K_{|\tilde{B}\beta|, |\tilde{\Lambda}\alpha|}$  according to (6.115).

Note that  $A$ ,  $B$ , and  $C$  are constant matrices independent of  $\alpha$  or  $\beta$ , and, thus, are precomputable.

**Example 5** (Delay estimation). We observe two signals, whereof the second signal has a variable time-delay as in Fig. 6.1. To estimate this time delay, we approximate the signals locally and at every index  $k$  by a fourth order polynomial and estimate the delay using (6.87). The result is likewise shown in Fig. 6.1.



**Figure 6.1:** Two-channel signal (CH1 and CH2) with variable time shift according to Example 5. The dashed line in CH2 is a copy of CH1 to illustrate the local time delay between the two channels. The true shift  $s$  and its local estimate  $\hat{s}$  after solving (6.87) are shown in the bottom plot.

◇

### Annex: Kronecker Product $\otimes$

The Kronecker product is defined, for any matrices  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{i \times j}$ , as

$$A \otimes B \triangleq \begin{bmatrix} A_{\{1,1\}}B & \cdots & A_{\{1,n\}}B \\ \vdots & & \vdots \\ A_{\{m,1\}}B & \cdots & A_{\{m,n\}}B \end{bmatrix} \quad (6.88)$$

with  $A_{\{m,n\}} \in \mathbb{R}$  the element in the  $m$ th row and  $n$ th column of  $A$  [28].

In the following, we give a brief summary of properties of Kronecker products, collected and modified from [28, 37].  $A$  and  $A'$  are both in  $\mathbb{R}^{m \times n}$ ,  $B$  and  $B'$  are both  $\in \mathbb{R}^{i \times j}$ , and  $C$  is of any finite dimension. Then we have

$$A^T \otimes B^T = (A \otimes B)^T \quad (6.89)$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) \quad (6.90)$$

$$C \otimes (A + A') = (C \otimes A) + (C \otimes A') \quad (6.91)$$

$$(A + A') \otimes C = (A \otimes C) + (A' \otimes C) \quad (6.92)$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD \quad (6.93)$$

$$\text{Tr}(A \otimes B) = \text{Tr}(A) \text{Tr}(B) \quad (6.94)$$

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B) \quad (6.95)$$

$$\text{vec}(AB) = (I \otimes A) \text{vec}(B) \quad (6.96)$$

$$\text{vec}(A) = (I_n^T \otimes A) \text{vec}(I_n) \quad (6.97)$$

$$\begin{aligned} \text{vec}(A \otimes B)^T \text{vec}(A' \otimes B') &= \\ & (\text{vec}(A) \otimes \text{vec}(B))^T (\text{vec}(A') \otimes \text{vec}(B')) \end{aligned} \quad (6.98)$$

$$\text{vec}(A \text{ diag}(c) B) = ((B^T \otimes 1_a) \circ (1_b \otimes A))c \quad (6.99)$$

$$\text{vec}(A \otimes B) = R_{m,n;i,j}(\text{vec}(A) \otimes \text{vec}(B)) \quad (6.100)$$

$$R_{m,n;i,j}^T \text{vec}(A \otimes B) = \text{vec}(A) \otimes \text{vec}(B) \quad (6.101)$$

with *permutation matrix*

$$R_{m,n;i,j} \triangleq I_n \otimes K_{m,j} \otimes I_i \in \mathbb{R}^{mni \times mni} \quad (6.102)$$

and for square  $A$  and  $B$ , i.e.,  $n = m$  and  $i = j$ ,

$$R_{m;i} \triangleq R_{m,m;i,i} \quad (6.103)$$

with  $K_{m,j}$  as in (6.115).

Applying the Kronecker product to vectors leads to further properties. Let  $a$  and  $b$  be column vectors of finite length, and  $a'$  a column vector of the same size as  $a$  and  $b'$  a column vector of the same size as  $b$ , respectively. Then we have

$$a \otimes b = \text{vec}(ba^\top) \quad (6.104)$$

$$a \otimes b = (I \otimes b)a \quad (6.105)$$

$$a^\top \otimes b = ba^\top = b \otimes a^\top \quad (6.106)$$

$$\text{diag}(b) = I_b \circ (1b \otimes b^\top) \quad (6.107)$$

$$= I_b \circ ((1^b)^\top \otimes b) \quad (6.108)$$

$$= I_b \circ (b \otimes (1^b)^\top) \quad (6.109)$$

$$= I_b \circ (b^\top \otimes 1^b) \quad (6.110)$$

$$(a \otimes b)^\top (a' \otimes b') = (b \otimes a)^\top (b' \otimes a') \quad (6.111)$$

$$(a \otimes b) = \text{vec}(ba^\top) = K_{|a|,|b|} \text{vec}(ab^\top) \quad (6.112)$$

$$= K_{|a|,|b|} (b \otimes a) . \quad (6.113)$$

The *commutation matrix*  $K_{m,n} \in \mathbb{R}^{mn \times mn}$  is a special case of the permutation matrix and commutes any matrix  $A \in \mathbb{R}^{m \times n}$ , i.e.,

$$K_{m,n} \text{vec}(A) = \text{vec}(A^\top) \in \mathbb{R}^{mn \times mn} \quad (6.114)$$

with

$$K_{m,n} = \sum_{i=1}^m \sum_{j=1}^n (e_{m,i} e_{n,j}^\top) \otimes (e_{n,j} e_{m,i}^\top) = (K_{n,m})^\top \in \mathbb{R}^{mn \times mn} \quad (6.115)$$

with the vectors  $e_{m,i} = [0, \dots, 0, 1, 0, \dots, 0]^\top$ , having a single 1 at the  $i$ th row [37].

### Annex: Hadamard Product (Elementwise Product) $\circ$

Let matrix  $A \in \mathbb{R}^{M \times N}$ , and vectors  $a, b \in \mathbb{R}^M$ , and  $c, d \in \mathbb{R}^N$ .

$$(a \circ b) \otimes (c \circ d) = (a \otimes c) \circ (b \otimes d) \quad (6.116)$$

$$a \otimes d = (a \otimes \mathbf{1}_N) \circ (\mathbf{1}_M \otimes d) \quad (6.117)$$

as a special case of (6.116)

$$a \circ b = \text{diag}(a)b = \text{diag}(b)a \quad (6.118)$$

$$= [I_M \circ (\mathbf{1}_M \otimes a^\top)]b \quad (\text{cf. (6.107)}) \quad (6.119)$$

$$A(a \circ b) = [A \circ (\mathbf{1}_M \otimes a^\top)]b \quad (6.120)$$

$$(Ab) \circ c = \text{diag}(c)Ab = [I_c \circ (\mathbf{1}^c \otimes c^\top)]Ab \quad (6.121)$$

$$= \text{diag}(Ab)c = [I_{Ab} \circ (\mathbf{1}^{Ab} \otimes (Ab)^\top)]c \quad (6.122)$$

### Annex: Trace

$$\text{Tr}(AB) = \text{vec}(A^\top)^\top \text{vec}(B) \quad (6.123)$$

$$\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B) \quad (6.124)$$

Part III

Applications and  
Examples





## Chapter 7

# Activation Map Reconstruction From EsoECG Signals

### 7.1 Introduction

As summarized in Section 3.3.2, a cardiac activation map provides a diagnostic tool in cardiology, visualizing the propagation of the electrical activity on the heart surface. In this chapter, we provide the details on how an activation map of the left atrial posterior wall is reconstructed, based on measurements from our esophageal catheter, EsoECG-3D.

As the solution to this problem is not unique (we only have a very limited number of observations while the propagation patterns on the heart surface are highly diverse), this problem is known to be “ill-posed” and denoted as the *inverse problem in electrocardiography* [34]. Consequently, an effective regularization on the result, or as we will do, the application of a restrictive model, is required.

### 7.2 Solving the Inverse Problem Using ALSSMs in a Filter Bank

One way to efficiently solve our inverse problem using our methods is through the application of multiple ALSSMs and to form a filter bank,

as we have proposed in [54]. Thereby, each ALSSM in the filter bank is parametrized such that it generates a signal corresponding to a physical model; in this case the physical model is a single moving electric dipole. Multiple such ALSSMs then form a filter bank to cover multiple measuring channels and multiple parameter selections.

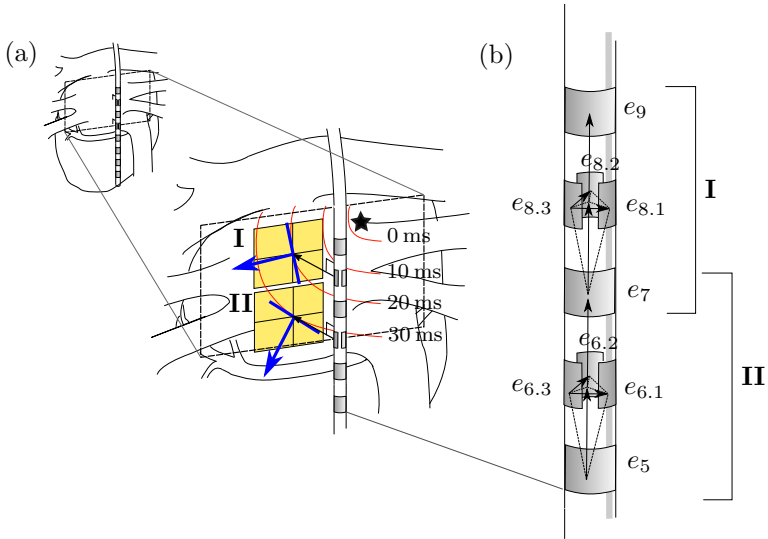
This method was applied to in vitro recordings using our 3-D catheter and has successfully localized single moving electric dipoles in 3-D space. Overall, this method is convincing in its efficiency, but it restricts the possible electrode arrangements on the catheter, which limits its use for our project.

## 7.3 Solving the Inverse Problem Using Polynomials

This second approach performs the computations in the domain of local polynomials, using the methods proposed in Chapter 6. Therefore, we first choose an appropriate model for the cardiac field. Then, we approximate the model's output signals by localized polynomials, as well as for the measured EsoECG signals. Finally, we directly match the two sets of polynomials, those originating from the models and those from the measurements, in their own feature space. According to Section 6.3.3, we here allow the polynomials to scale in amplitude and time (dilation). Note that matching the signals with variable time dilation is essential, since we do not know the physical propagation speed of the observed cardiac activations and, thus, also not the temporal spreading of their electrical signals.

### 7.3.1 Method

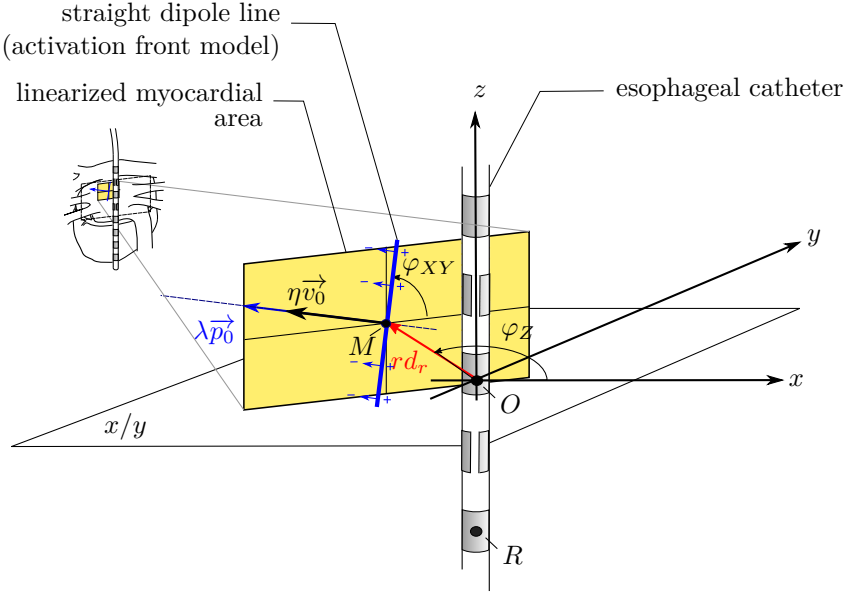
We recall that we want to estimate the spatial propagation of the cardiac activation front on the heart surface based on the measured signals. For that, we need an effective regularization which is, in this approach, introduced by a restrictive model. We here use the model of a straight moving dipole line in 3-D space to approximate the activation front on the heart surface; we denote a straight dipole line a dipole, where its dipole moment is spread homogeneously along a straight line of finite length as depicted in Fig. 7.1 and Fig. 7.2. We use this model to match the activation front locally, i.e., only on a small area on the heart sur-



**Figure 7.1:** Modeling the myocardial activation process by multiple dipole lines after subdividing the atrial surface into multiple areas (yellow shaded squares). Figure (a): Left atrial posterior wall with electrical trigger point  $\star$  and the observed activation front at time 0 ms, 10 ms, 20 ms, and 30 ms (thin red curves). Straight dipole line models (thick blue lines) with their local observation areas, propagation directions (blue arrows), and the assigned sets of electrodes **I** and **II**. Figure (b): Zoomed catheter view with labeled electrodes ( $e_5, e_{6.1}, \dots$ ), the measured potential differences (black arrows), and the assignment to the sets of electrodes **I** and **II**.

face, where we assume the movements to be linear and to propagate at constant speed.

For that, we first subdivide the myocardium of the left atrial posterior wall into multiple, possibly overlapping areas, which are processed independently. Then we assign to each such area those electrodes which are close; the minimum number of electrodes required per area depends on the complexity of the model to be estimated. For our model with only 5 parameters, 5 electrodes per area are sufficient. An example of 2 such areas of approximately  $1.50 \text{ cm} \times 1.50 \text{ cm}$  in size, along with their assigned set of electrodes is depicted in Fig. 7.1. Note that certain elec-



**Figure 7.2:** Parametrization of the straight dipole line model and the orthogonal reference coordinate system  $x/y/z$  (cf., Table 7.1). Additionally shown are: dipole line (solid blue line) and myocardium area to be analyzed (yellow area).  $O$  denotes the origin of the coordinate system and  $R$  a center point of one of the electrodes.

trodes can belong to more than one such set at the same time. The final activation map is then an aggregation of the results of the single areas<sup>1</sup>.

### 7.3.2 Physical Model

To use the physical model of a straight dipole line, we need to introduce its parameters and to map them to our specific application.

#### Straight Dipole Line Model

Let  $M$  be the closest point on a straight dipole line to a center point  $O$  on the catheter, both in 3-D space (see Fig. 7.2). Furthermore,  $r \in \mathbb{R}_+$

<sup>1</sup>We are aware that the results could be improved, if the inversion for all areas was performed jointly. This idea should be considered in a later version.

**Table 7.1:** Dipole line model parametrization according to Fig. 7.2 after reducing the number of model parameters.

| Symbol                         | Description  | Unit             |
|--------------------------------|--|------------------|
| $\lambda \in \mathbb{R}$       | dipole line strength per line unit   | mAs/m            |
| $\eta \in \mathbb{R}_+$        | activation front speed   | $\text{ms}^{-1}$ |
| $r \in \mathbb{R}_+$           | minimal activation front distance to the catheter (orthogonal to the myocardium plane) | m                |
| $\varphi_{XY} \in [-\pi, \pi]$ | activation front angle   | rad              |
| $\varphi_Z \in [-\pi, \pi]$    | catheter rotation angle  | rad              |

denotes the distance from  $M$  to  $O$  and  $\vec{d}_r \in \mathbb{R}^3$  the unit direction of the position vector  $\overline{OM}$ ,  $\lambda \in \mathbb{R}$  the dipole moment with unit direction  $\vec{p}_0 \in \mathbb{R}^3$ , and  $\eta \in \mathbb{R}$  the dipole line movement speed with its unit movement direction  $\vec{v}_0 \in \mathbb{R}^3$ . Note that we will assume  $\vec{p}_0$  and  $\vec{v}_0$  to be parallel, as we will see later.

The electrical field strength produced by a straight moving dipole line and observed at point  $R$  is

$$\phi_R(t) = \int_a^b \rho \frac{\langle \eta \vec{p}_0, (r \vec{d}_r + \eta \vec{v}_0 t + x \vec{l} - \overline{OR}) \rangle}{\|r \vec{d}_r + \eta \vec{v}_0 t + x \vec{l} - \overline{OR}\|^3} dx \in \mathbb{R} \quad (7.1)$$

with  $\vec{l} \in \mathbb{R}^3$  the unit vector pointing along the dipole line,  $a \in \mathbb{R}$  and  $b \in \mathbb{R}$  the length of the dipole line to both sides of the center point  $M$ , and  $\rho \in \mathbb{R}$  some constant value, summarizing all physical constants. In our case,  $R$  is chosen as the location of one of the catheter's electrodes. Note that the field in (7.1) is derived in analogy to [54].

### Parameter Reduction and Reparametrization

We note that (7.1) has more than 10 scalar parameters, which all need to be specified. Fortunately, the number of parameters reduces considerably if we take anatomical and physiological considerations into account. First, we use a fact known from experience in physiology: the directions of the dipole movement and of the dipole moment are in good approximation parallel [32, Chapter 5.3], i.e., we assume that  $\vec{p}_0 \approx \vec{v}_0$ . Second,

considering the human body’s anatomy, we see that the left atrium posterior wall is roughly in parallel to the catheter axis, i.e.,  $\langle \vec{d}_r, \vec{c}_0 \rangle \approx 0$ . To apply these dependencies, we introduce the new angles  $\varphi_{XY}$  and  $\varphi_Z$ , indicating the angle between the x/y plane of our coordinate system and the dipole line (cf. Fig. 7.2). Reparametrizing the model accordingly reduces the number of parameters to 5 scalars, which are listed in Table 7.1.

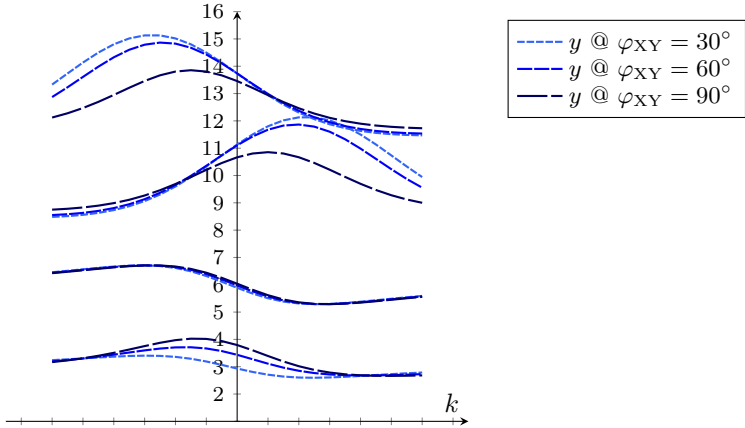
### 7.3.3 Model Inversion

To estimate the two model parameters dipole strength  $\lambda$  and front speed  $\eta$ , we have closed forms, as we will see later. To handle the remaining non-linear parameters, we discretize them according to their physiological range, as in Table 7.2. Then, we generate with our physical model for each selection of parameters a synthetic multi-channel EsoECG output signal, and substitute these outputs by local polynomials. We subsequently denote such a multi-channel polynomial approximation a *template*. An example of parameter discretization is given in Table 7.2, and polynomials approximating the outputs of the model are shown in Fig. 7.3.

We equally approximate the EsoECG measurements by local polynomials, and then match these local polynomials to all templates; the polynomials are matched directly in their feature space, for which we have derived a closed form solution in Section 6.3.3.

**Table 7.2:** Example of model parameter discretization. Note that the dipole strength  $\lambda$  and front speed  $\eta$  are estimated in closed form and are, therefore, not discretized.

| Symbol         | Discretization                | #    |
|----------------|-------------------------------|------|
| $\lambda$      | (closed form)                 | -    |
| $\eta$         | (closed form)                 | -    |
| $r$            | $1 \times 10^{-2}$ m          | 1    |
| $\varphi_{XY}$ | $[0, 1, 2, \dots, 35] \pi/18$ | 36   |
| $\varphi_Z$    | $[0, 1, 2, \dots, 35] \pi/18$ | 36   |
|                | Total                         | 1296 |



**Figure 7.3:** Examples of local polynomial approximations of synthetic 4-channel EsoECG signals. The synthetic signals were generated using the physical model of a moving dipole line with parameters:  $r = 10$  mm,  $v = 1$  m/s,  $\varphi_Z = 30^\circ$ ,  $\varphi_{XY} = 30^\circ / 60^\circ / 90^\circ$ , and  $\lambda = 1$  mAs/m. To improve readability, an individual offset of 3 to 12 mV was added to each channel.

### Template Matching With Variable Time Dilation

We note that a decrease in the activation front speed  $\eta$  leads to a dilation in time of the observed signal shapes; if we know the signal dilation, we also know the source's speed.

Let  $\alpha_r \in \mathbb{R}^{(Q+1) \times M}$  be the polynomial coefficients of such an  $M$ -channel template,  $r$  denotes the template index and  $Q$  the polynomial order. Furthermore,  $\beta_k \in \mathbb{R}^{(Q+1) \times M}$  are the coefficients of the local polynomial approximation of our measured EsoECG signal around time index  $k$ .

Due to unknown baseline offsets superimposed on all signal channels, we use the offset-free comparison of Section 6.3.3 to estimate the signal amplitude  $\lambda_r$  and the signal propagation speed  $\eta_r$  independent of any per-channel offset voltage: according to (6.78), the speed estimate is

$$\hat{\eta}_r = \operatorname{argmin}_{\eta} \frac{(\tilde{\alpha}_{r,2} \otimes \tilde{\alpha}_{r,2})^\top \eta^{Bq}}{\tilde{\alpha}_{r,3}^\top, \eta^{Mq}}, \quad (7.2)$$

with  $\tilde{\alpha}_{r,2}$  and  $\tilde{\alpha}_{r,3}$  according to (6.73) and (6.74), and after substituting  $\beta_k$  by  $\beta'_k$  and  $\alpha_r$  by  $\alpha'_r$  as in (6.85). The estimate of the dipole moment scalar  $\hat{\lambda}_r$  is according to (6.76)

$$\hat{\lambda}_r = \underset{\lambda}{\operatorname{argmin}} J_r(\lambda, \hat{\eta}_r) = -\frac{\tilde{\alpha}_{r,2}^T \hat{\eta}_r^q}{\tilde{\alpha}_{r,3}^T \hat{\eta}_r^{Mq}} . \quad (7.3)$$

Finally, after having fit each individual template, we take that template  $\hat{r}$  with the smallest remaining square error, which is

$$\hat{r} = \underset{r}{\operatorname{argmin}} J_r(\hat{\lambda}_r, \hat{\eta}_r) . \quad (7.4)$$

Note that the least squared error  $J_r(\cdot)$  when fitting template  $r$  to our signal at index  $k$ , conveniently follows from (6.71) and is

$$J_r(\hat{\lambda}_r, \hat{\eta}_r) = \tilde{\alpha}_{r,1} - 2\hat{\lambda}_r \tilde{\alpha}_{r,2}^T \hat{\eta}_r^q + \lambda_r^2 \tilde{\alpha}_{r,3}^T \hat{\eta}_r^{Mq} . \quad (7.5)$$

Figure 7.4 shows an example of an atrial EsoECG signal of approximately 30 ms duration for a set of 4 electrodes and depicts the local fit after successfully optimizing the model parameters and selecting the best matching template. Figure 7.5 shows a longer EsoECG sequence and additionally displays the estimated model parameters for each detected atrial signal wave. First graphical results of aggregated activation maps, i.e., after aggregating multiple results derived using different sets of electrodes, were already given in Fig. 3.6, Section 3.3.2.

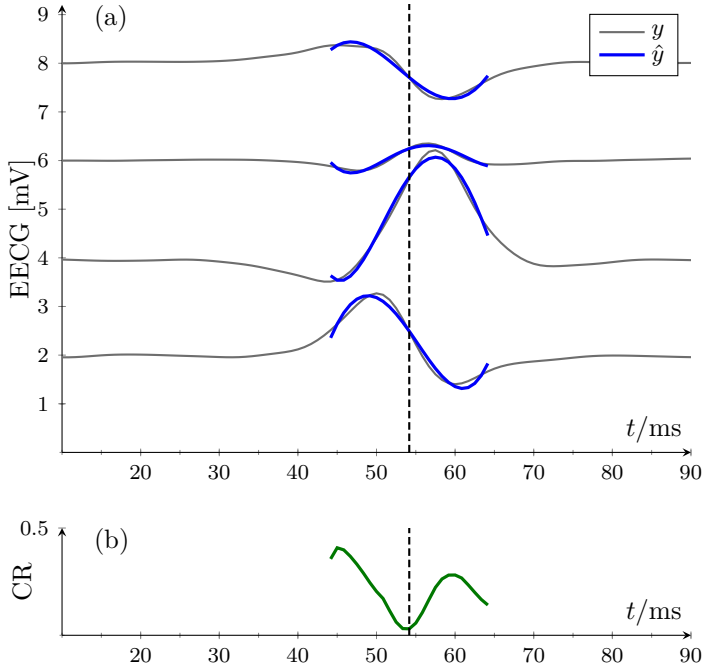
## Detection of Actual Wavefronts

At this point, the discrimination between real wave fronts and signal artifacts is still missing; we so far applied our model blindly at any time index  $k$ , regardless of whether at that moment a wave front passed by or not. To make up for this, we will use the measure of the cost ratio (4.80) from Section 4.6.

We recall that the model's distance vector  $\vec{d}_r$  was defined to be perpendicular to the activation front movement direction  $\vec{v}_0$ . But clearly this can only be true for a single instant as our activation fronts are moving. As we do not know this instant a priori, we first applied our model at every time index  $k$ , and now use the cost ratio (4.80) retroactively: we are looking for a local minimum in

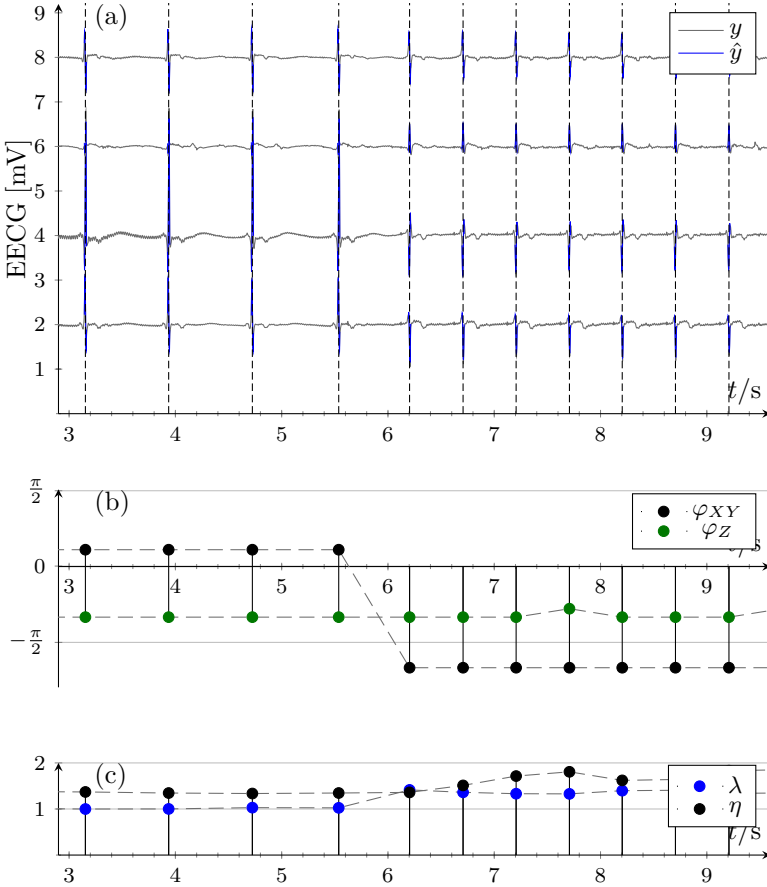
$$\text{CR} = \frac{J(\lambda_{\hat{r}}, \eta_{\hat{r}})}{J(0, \eta_{\hat{r}})} . \quad (7.6)$$





**Figure 7.4:** Example of parametrized dipole line model fitting a multi-channel EsoECG signal sequence. Plot (a) shows a 4-channel EsoECG signal  $y$  of an atrial beat. The model output  $\hat{y}$  is generated using best matching parameters with index  $\hat{r} = 423$  ( $\varphi_{XY} = -30^\circ$ ,  $\varphi_Z = -70^\circ$ ) and time dilation  $\eta = 1.12$  and amplitude scaling  $\lambda = 1.43$ . These model parameters were selected at the local minimum of the CR in plot (b) at  $t = 54$  ms (dashed line), as in (7.6) on page 94. To improve readability, an individual offset of 2 to 8 mV was added to each channel in Plot (a). Data source: clinical trial *esoECG-3D* patient IDA07 around  $t = 5732.97$  s.

Figure 7.4 depicts this CR over a short time period containing a single activation front, indicating a clear local minimum. Figure 7.5 (a) depicts an EsoECG of multiple heart beats, where we detect in each beat the activation front of the atrial activity, and Fig. 7.5 (b) and (c) provide the corresponding model parameters estimated for each detected activation front.



**Figure 7.5:** Model parameter estimation for atrial waves in an EsoECG signal. Plot (a) shows a 4-channel EsoECG signal of 10 seconds duration. At time  $t = 6$  s, electric stimulation starts using an intra-cardiac pace maker. Plots (b) and (c) show the estimated model parameters (Table 7.1) at the location of minimal CR (7.6).  $\varphi_{XY}$  changes for  $t > 6$  s due to the stimulations, from  $20^\circ$  to  $-120^\circ$ , indicating an alteration of the trigger's origin.  $\varphi_Z$  is not affected by the stimulations; this parameter is the rotation of the catheter with respect to the heart. Fig. 7.4 shows a zoomed version of this figure. To improve readability, an individual offset of 2 to 8 mV was added to each channel in Plot (a). Data source: clinical trial *esoECG-3D* patient IDA07 at  $t_0 = 5559$  s.

## Chapter 8

# Esophageal Isopotential Map Reconstruction (Supplements)

Esophageal isopotential maps (EsoIPMs), as introduced in Section 3.3.1, have already been published by us in 2018 [50]. We here supplement this published method by a closed-form solution using our calculus for polynomials from Chapter 6. Note that our calculus was not yet available at the time of that publication.

### **Katheter Depth Estimation**

In [50] we introduced univariate polynomials of order  $N_\varphi$  in  $z \in \mathbb{R}$ , denoted as  $\varphi_j(z) \in \mathbb{R}$ , interpolating the esophageal measurements across all measuring channels at a single time index  $j \in \mathbb{N}$ ;  $z$  is the position along the catheter (and the esophagus).

As we have learned, such an esophageal catheter is constantly drifting back and forth due to breathing and body motions and the like. To obtain an estimate of that vertical catheter shift between two heart beats, we compare the EsoECG signals recorded at corresponding time indices  $j$  and  $j'$ . We need to find the shift  $\hat{s}$ , minimizing the squared error between the two polynomials  $\varphi_j(z - \hat{s}/2)$  and  $\varphi_{j'}(z + \hat{s}/2)$  for a given interval of  $z$  [50, Equation (1)]. Note that we here shift both polynomials

symmetrically by  $\hat{s}/2$  such that the interval center point remains at  $z$ . We will solve this problem using our calculus.<sup>1</sup>

For that, we first assign  $\varphi_j(z) \hat{=} \alpha^\top z^q$  and  $\varphi_{j'}(z) \hat{=} \beta^\top z^q$  with coefficient vectors  $\alpha \in \mathbb{R}^N$  and  $\beta \in \mathbb{R}^N$ , and exponent vectors  $q = [0, \dots, N_\varphi]^\top$  and get the cost function

$$J(s) = \int_a^b \left[ \alpha^\top \left( z - \frac{s}{2} \right)^q - \beta^\top \left( z + \frac{s}{2} \right)^q \right]^2 dz . \quad (8.1)$$

This cost function is of the same form as the already solved problem (6.86) in Section 6.3.4 and is therefore minimized by

$$\hat{s} = \underset{s}{\operatorname{argmin}} \tilde{\alpha}^\top s^{q'} \quad (8.2)$$

with

$$\tilde{\alpha} = (A(\alpha \otimes \alpha) - B(\alpha \otimes \beta) + C(\beta \otimes \beta)) , \quad (8.3)$$

and  $A$ ,  $B$ ,  $C$ , and  $q'$  according to (6.87).

We note that  $A$ ,  $B$ , and  $C$  are independent of the data samples which are incorporated in  $\alpha$  and  $\beta$ . Therefore, the repeated minimization of (8.1) for multiple data samples is efficiently executed.

---

<sup>1</sup>In contrast to the published version, we here use fixed and symmetric integration boundaries, as they simplify the computations and we expect the solution to be more robust.

# Chapter 9

## Windowed State Space Filters (Summary)

We here briefly introduce our framework “*Windowed State-Space Filters for Signal Detection and Separation*” [47] based on ALSSMs for signal detection and signal filtering tasks. This framework was published in 2018 and is an immediate application of Chapter 4.<sup>1</sup>

This framework introduces a uniform formal and graphical notation (cf. Fig. 9.1) to design filters for practical applications, using the cost functions derived in Section 4. Cost terms based on a single ALSSM supplemented by its window are denoted as *cost segments*, while the combination of multiple such cost segments, leading to a multi-segment window, is denoted as a *composite cost*. Finally, it also takes use of time-variant windows as introduced in Section 4.2.5.

This framework provides a ready-to-use method and includes all relevant recursive computations in a tabular overview, supplemented by multiple examples.

### The Cost Segments

We denote a *cost segment* the quadratic cost function

$$J_a^b(k, x, \theta) = \sum_{i=k+a}^{k+b} \alpha_{k+\delta}(i) v_i (y_i - cA^{i-k}x)^2 \quad (9.1)$$

---

<sup>1</sup>Note that [47] is written *before* the method of *window* ALSSMs was derived. Thus, this method is lacking in the published framework.

over a fixed interval  $\{a, \dots, b\}$ , with  $a \in \mathbb{Z} \cup \{-\infty\}$ ,  $b \in \mathbb{Z} \cup \{+\infty\}$ , and  $a \leq b$ , and with initial state vector  $x \in \mathbb{R}^{N \times 1}$  and parameter set  $\theta = (c, \delta, \alpha)$  where  $c \in \mathbb{R}^{1 \times N}$ ,  $\delta \in \mathbb{Z}$  and  $\alpha = [\alpha_k(i)]_{(k,i) \in \mathbb{Z}^2}$ . This cost is of a similar form as (4.57) on page 39 and also computed recursively. Note that the shift  $\delta$  introduced in (9.1) leads to the normalization of the window to a value of 1 at index  $\delta$  and thereby simplifies the window design. This is an immediate application of the property  $\alpha_{k+\delta}(k+\delta) = 1$  for all  $(k, \delta) \in \mathbb{Z}^2$ , derived from (4.59). Finally,  $v_i \in \mathbb{R}$  is the per-sample weight as introduced in (4.48) on page 38.

### The Composite Cost

The cost segment (9.1) is nicely computable, but it is restricted to a single model  $(c, A)$  and a single window. To overcome these limitations, we join several such cost segments together, as proposed in Section 4.2.6 on page 40, and create a more versatile cost, called a *composite cost*.

We define  $P \in \mathbb{N}$  cost segments with parameters  $\theta_p = (c_p, \delta^{(p)}, \alpha^{(p)})$  and boundaries  $a_p \in \mathbb{Z} \cup \{-\infty\}$ ,  $b_p \in \mathbb{Z} \cup \{+\infty\}$ ,  $a_p \leq b_p$ ,  $p \in \{1, \dots, P\}$ , leading to the general form of a *composite cost*

$$\tilde{J}(k, x, \Theta) = \sum_{p=1}^P \beta_p J_{a_p}^{b_p}(k, x, \theta_p), \quad (9.2)$$

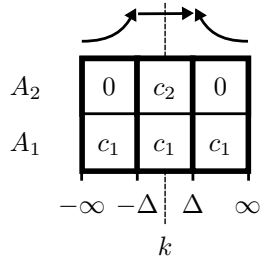
with  $\Theta = (\theta_1, \theta_2, \dots, \theta_P)$  and the *segment scalars*  $\beta_p \in \mathbb{R}_+$ ; the segment scalar  $\beta_p$  is an engineering parameter to weight cost terms individually.

### A Graphical Notation

To efficiently design new such composite costs for particular applications, we introduce the graphical notation as shown in Fig. 9.1. This notation arranges the used models in a grid, indicating time on the horizontal axis and superposition of models on the vertical. In this grid, each column denotes a cost segment and each row a single ALSSM model; multiple models in a single column indicates model superposition (see *Linear combination of M systems* in Table 4.1 on page 30). At the top of each column, an arrow further indicates the applied window shape and the direction of stable recursion computations.

### Example: Separation of Pulse and Baseline Interference

This is a basic example to separate a pulse from its baseline interferences. Figure 9.3 shows a noisy single-channel EsoECG signal with two



**Figure 9.1:** Graphical representation of a composite cost. Each column denotes a cost segment and each row a single ALSSM model. This cost is composed of two superimposed models  $(A_1, c_1)$  and  $(A_2, c_2)$ : while the first model contributes with its non-zero output vector  $c_1$  to all three segments, the second model only contributes with  $c_2$  to the middle segment in the time interval  $\{-\Delta, \dots, \Delta - 1\}$ , but has zero outputs in the other two segments. At top of each column, an arrow indicates the applied window shape (exponentially decaying in the first and the last segment, and rectangular in the middle segment), and the computation direction of stable recursions.

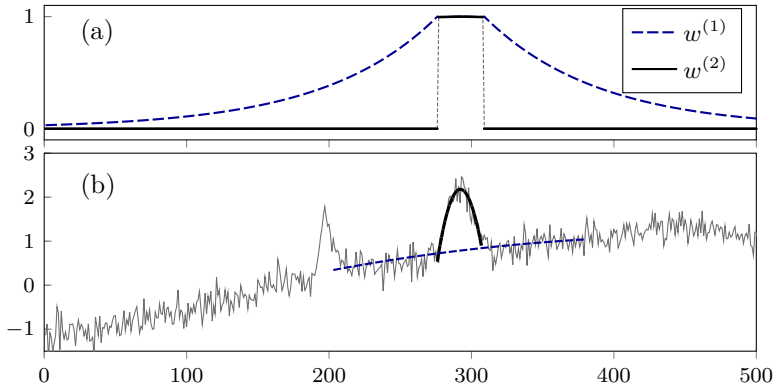
pulses superimposed by a baseline signal. To locally decompose the signal, we apply the composite cost as in Fig. 9.1 with 3 segments and 2 models, where model  $(A_1, c_1)$ , a 2nd order polynomial, generates the baseline over a two-sided decaying window, and model  $(A_2, c_2)$ , a 3rd order polynomial, the pulse signal with a rectangular window of finite support. Finally, to identify the pulse locations and separate them from the baseline, we look for local maxima in the cost ratio (4.82)

$$\text{LCR}_k = -\frac{1}{2} \log \frac{\tilde{J}(k, [\hat{x}_2], \Theta)}{\tilde{J}(k, [\hat{x}_1], \Theta)} \quad (9.3)$$

with

$$\{\hat{x}_1, \hat{x}_2\} = \underset{x_1, x_2}{\operatorname{argmin}} \tilde{J}(k, [x_2], \Theta). \quad (9.4)$$

The pulse estimates along with the local baseline estimates at a single  $\text{LCR}_k$  maximum are shown in Fig. 9.3.

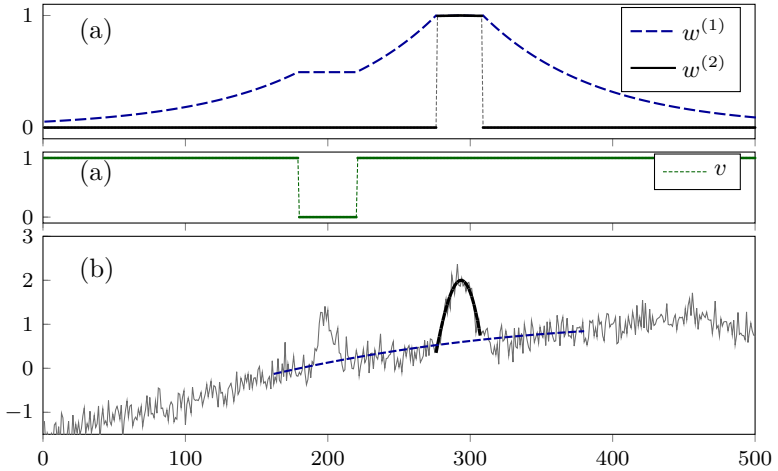


**Figure 9.2:** Local decomposition of an ECG signal wave from its base-line signal. This local decomposition applies the composite cost as in Fig. 9.1 with 3 segments and 2 models. Plot (a) shows the two-sided exponentially decaying window  $w^{(1)}$ , a composition of all three window segments. This composed window applies to the first model ( $A_1, c_1$ ) only and is fitting the baseline interferences. The second window  $w^{(2)}$  is of finite support and only active for the middle segment; this window applies to the second model ( $A_2, c_2$ ) and fits the ECG pulses. Plot (b) shows the model trajectories after minimization of the composite cost: the trajectory of the first model (blue dashed line) corresponds to that of the local baseline estimate, and the trajectory of the second model (black line) corresponds to the estimate of the pulse shape.

### Example: Separation of Pulse and Baseline Interference Using a Time-Variant Window

We frequently process signals in multiple consecutive steps, where intermediate results are supplemented with some per-sample reliability measure or sample weight. In this example, we incorporate this prior knowledge and weight each sample in our squared cost function accordingly. Furthermore, we use the method of time-variant windows, cf. Section 4.2.5, to keep the weighted number of samples within each squared error function constant; this elongates the window wherever samples of low or zero sample weights are involved, and narrows the window whenever the sample weights are elevated. Figure 9.3 shows the same noisy single-channel EsoECG signal as already shown in Fig. 9.2, but now applying a time-variant window using the given sample weights.





**Figure 9.3:** Local decomposition of an ECG signal wave from its baseline signal as in Fig. 9.2, but using a time-variant windows. Plot (a) shows the two-sided exponential decaying window  $w^{(1)}$ , now with a plateau whenever the samples are weighted zero, i.e.,  $v_k = 0$ . Plot (b) shows the a priori assigned sample weight  $v_k \in [0, 1]$  for each sample. Plot (c) shows the model trajectories as in Fig. 9.3, but with an extended window to the left and a improved baseline fit due to applied sample weights.



## Chapter 10

# Further Applications and Examples

We here provide a (non-exhaustive) list of further examples and applications, which are based on our methods and already published.

## 10.1 Examples on Signal Detection and Discrimination Using ALSSMs

**Sinusoidal Pulse Detection.** Robust detection of a single pulse of sinusoidal shape in a noisy single-channel signal. See [47, Paragraph V. A].

**Separation of Unknown Pulses From Baseline Signal.** This example demonstrates the separation of a pulse of (approximately) known duration but of unknown shape from a interference or baseline signal. See [47, Section V. B].

**ECG Signal Baseline Recovery.** ECG signals, such as most biological signals, are superimposed by an interference signal, denoted as a baseline signal. This example demonstrates how to separate the ECG from its baseline signal. See [47, Section V. B].

**Morse Code Detection.** Morse codes are an excellent textbook example to demonstrate the robust detection of a known signal within noisy observations. This example demonstrates the use of multi-segment windows (cf. 4.2.6) to generate a model for a Morse code symbol. See [47, Section V. F].

**Detection of a Modulated Signal.** This example demonstrates the multiplicability of ALSSM: The product of two ALSSMs is used to model and detect an amplitude modulated signal. See [55, Section V. A.].

**ECG Wave Discrimination Using Multi-Timescale Baseline Estimates.** In this example, ALSSMs and local polynomial approximations are used to efficiently detect and discriminate atrial and ventricular ECG waves of distinct shapes in a single-channel esophageal ECG signal. The ALSSMs are composed of multiple models and not only approximate the ECG waves but also the baseline signal. See [48, Section 4.2].

**P Wave Discrimination in Multi-Channel ECG.** This example demonstrates the use of ALSSMs in a multi-channel ECG signal. The ECG signal used is of low SNR and the P waves of interest only vary very little in shape. Here the multi-channel approach and the use of fixed-amplitude signal templates help to increase the detection and discrimination robustness. See [48, Section 4.3].

**Model Switch Detection - A Toy Example.** A further basic example of model switch detection, along with an in-detail description. See [47, Section III.G].

**Outlier Detection - A Toy Example.** A further basic example of outlier detection, along with an in-detail description. See [47, Section III. G].

## 10.2 Examples on ALSSMs in a Filter Bank

**Inferring Depolarization of Cells from 3D-Electrode Measurements Using a Bank of Linear State Space Models.** Solving the *inverse problem* as introduced in Section 7 is computationally intense. In this example, a bank of ALSSMs is used to efficiently address a simplistic form of this problem. This method was tested on basic in-vitro measurements. See [54].

## 10.3 Examples on Signal Interpolation and Reconstruction Using ALSSMs

**Reconstruction of a Signal From Its Quantized Version.** Quantizing a signal's amplitude drops the signal's smoothness. In this example, a localized polynomial in combination with window of adaptive length is used to reconstruct a smooth version of the unquantized signal. See [47, Section V.E].

## 10.4 Applications of Polynomial Cost Functions

**Estimation of the Cardiac Field in the Esophagus Using a Multipolar Esophageal Catheter.** This journal article demonstrates the use of polynomial cost functions in order to solve the catheter tracking problem introduced in Section 3.3.1. See [50].

# Appendices





# Appendix A

## Derivations in Chapter 4

### A.1 Derivation of (4.6)

To find the parametrized form from the cost term (A.1), we write

$$\begin{aligned}
 & \sum_{i=k+a}^{k+b} w_{i-k} (cA^{i-k}x - y_i)^2 \tag{A.1} \\
 &= \sum_{i=k+a}^{k+b} w_{i-k} (cA^{i-k}x - y_i)^\top (cA^{i-k}x - y_i) \\
 &= \sum_{i=k+a}^{k+b} w_{i-k} x^\top (A^{i-k})^\top c^\top c A^{i-k} x \\
 &\quad - 2 \sum_{i=k+a}^{k+b} w_{i-k} x^\top (A^{i-k})^\top c^\top y_i + \sum_{i=k+a}^{k+b} w_{i-k} y_i^2 \\
 &= x^\top \left( \underbrace{\sum_{i=k+a}^{k+b} w_{i-k} (A^{i-k})^\top c^\top c A^{i-k}}_{W_k} \right) x - 2x^\top \left( \underbrace{\sum_{i=k+a}^{k+b} w_{i-k} (A^{i-k})^\top c^\top y_i}_{\xi_k} \right) \\
 &\quad + \underbrace{\sum_{i=k+a}^{k+b} w_{i-k} y_i^2}_{\kappa_k} \\
 &= x^\top W_k x - 2x^\top \xi + \kappa_k \tag{A.2}
 \end{aligned}$$

with

$$W_k \triangleq \sum_{i=k+a}^{k+b} w_{i-k} (A^{i-k})^\top c^\top c A^{i-k} \quad (\text{A.3})$$

$$\xi_k \triangleq \sum_{i=k+a}^{k+b} w_{i-k} (A^{i-k})^\top c^\top y_i \quad (\text{A.4})$$

$$\kappa_k \triangleq \sum_{i=k+a}^{k+b} w_{i-k} y_i^2. \quad (\text{A.5})$$

## A.2 Derivation of (4.28) and (4.29)

To derive (4.28), we write

$$\begin{aligned} W_k &= \sum_{i=k+a}^{k+b} ((A^{i-k})^\top c^\top c A^{i-k}) (\check{c} \check{A}^{i-k} \check{x}_0) \\ &= \sum_{i=k+a}^{k+b} ((A^{i-k})^\top c^\top) ((c A^{i-k}) \otimes (\check{c} \check{A}^{i-k} \check{x}_0)) \\ &= \sum_{i=k+a}^{k+b} ((A^{i-k})^\top c^\top) ((c \otimes \check{c}) (A^{i-k} \otimes (\check{A}^{i-k} \check{x}_0))) \\ &= \sum_{i=k+a}^{k+b} ((A^{i-k})^\top c^\top) ((c \otimes \check{c}) ((A^{i-k} I_N) \otimes (\check{A}^{i-k} \check{x}_0))) \\ &= \sum_{i=k+a}^{k+b} ((A^{i-k})^\top c^\top) ((c \otimes \check{c}) ((A^{i-k} \otimes \check{A}^{i-k}) (I_N \otimes \check{x}_0))) \\ &= \underbrace{\left( \sum_{i=k+a}^{k+b} (A^{i-k})^\top c^\top \widehat{c} \widehat{A}^{i-k} \right)}_{\widehat{W}_k} \underbrace{(I_N \otimes \check{x}_0)}_R \\ &= \widehat{W}_k R \end{aligned} \quad (\text{A.6})$$

where we use property (6.93) of Kronecker products in the third and fifth derivation step.

To derive (4.29), we write

$$\begin{aligned}
\xi_k^\top &= \sum_{i=k+a}^{k+b} (cA^{i-k})(\check{c}\check{A}^{i-k}\check{x}_0)y_i \\
&= \sum_{i=k+a}^{k+b} (c \otimes \check{c})(A^{i-k} \otimes (\check{A}^{i-k}\check{x}_0))y_i \\
&= \sum_{i=k+a}^{k+b} (c \otimes \check{c})((A^{i-k}I_N) \otimes (\check{A}^{i-k}y_i\check{x}_0)) \\
&= \left( \sum_{i=k+a}^{k+b} (c \otimes \check{c})(A^{i-k} \otimes \check{A}^{i-k})y_i \right) \underbrace{(I_N \otimes \check{x}_0)}_R \\
&= \underbrace{\left( \sum_{i=k+a}^{k+b} \widehat{c}\widehat{A}^{i-k}y_i \right)}_{\widehat{\xi}_k^\top} R \\
&= \widehat{\xi}_k^\top R
\end{aligned} \tag{A.7}$$

also using property (6.105) of Kronecker products.

To derive (4.30), we write

$$\begin{aligned}
\kappa_k &= \sum_{i=k+a}^{k+b} (\check{c}\check{A}^{i-k}\check{x}_0)y_i^2 = \left( \sum_{i=k+a}^{k+b} \check{c}\check{A}^{i-k}y_i^2 \right) \check{x}_0 \\
&= \underbrace{\left( \sum_{i=k+a}^{k+b} \check{c}\check{A}^{i-k}y_i^2 \right)}_{\widehat{\kappa}_k} \check{x}_0 \\
&= \widehat{\kappa}_k \check{x}_0,
\end{aligned} \tag{A.8}$$

### A.3 Derivation of (4.33), (4.33), and (4.34)

The forward recursions ( $k \rightarrow k+1$ ) for (4.33), (4.33), and (4.34) are

$$\begin{aligned}
\widehat{W}_{k+1} &= \sum_{i=k+a+1}^{k+b+1} (A^{i-k-1})^\top c^\top \widehat{c}\widehat{A}^{i-k-1} \\
&= A^{-\top} \left( \sum_{i=k+a}^{k+b} (A^{i-k})^\top c^\top \widehat{c}\widehat{A}^{i-k} \right) \widehat{A}^{-1} \\
&\quad - (A^{a-1})^\top c^\top \widehat{c}\widehat{A}^{a-1} + (A^b)^\top c^\top \widehat{c}\widehat{A}^b
\end{aligned}$$

$$= A^{-\top} \tilde{W}_k \hat{A}^{-1} - (A^{a-1})^\top c^\top \hat{c} \hat{A}^{a-1} + (A^b)^\top c^\top \hat{c} \hat{A}^b \quad (\text{A.9})$$

$$\begin{aligned} \hat{\xi}_{k+1} &= \sum_{i=k+a+1}^{k+b+1} \hat{c} \hat{A}^{i-k-1} y_i \\ &= \left( \sum_{i=k+a}^{k+b} \hat{c} \hat{A}^{i-k} y_i \right) \hat{A}^{-1} - \hat{c} \hat{A}^{a-1} y_{k+a} + \hat{c} \hat{A}^b y_{k+b+1} \\ &= \hat{\xi}_k^\top \hat{A}^{-1} - \hat{c} \hat{A}^{a-1} y_{k+a} + \hat{c} \hat{A}^b y_{k+b+1} \end{aligned} \quad (\text{A.10})$$

$$\begin{aligned} \hat{\kappa}_{k+1} &= \sum_{i=k+a+1}^{k+b+1} \check{c} \check{A}^{i-k-1} y_i^2 \\ &= \left( \sum_{i=k+a}^{k+b} \check{c} \check{A}^{i-k} y_i^2 \right) \check{A}^{-1} - \check{c} \check{A}^{a-1} y_{k+a}^2 + \check{c} \check{A}^b y_{k+b+1}^2 \\ &= \hat{\kappa}_k \check{A}^{-1} - \check{c} \check{A}^{a-1} y_{k+a}^2 + \check{c} \check{A}^b y_{k+b+1}^2 \end{aligned} \quad (\text{A.11})$$

## A.4 Derivation of (4.35), (4.37), and (4.38)

The backward recursions ( $k \rightarrow k-1$ ) for (4.35), (4.37), and (4.38) are

$$\begin{aligned} \hat{W}_{k-1} &= \sum_{i=k+a-1}^{k+b-1} (A^{i-k+1})^\top c^\top \hat{c} \hat{A}^{i-k+1} \\ &= A^\top \left( \sum_{i=k+a}^{k+b} (A^{i-k})^\top c^\top \hat{c} \hat{A}^{i-k} \right) \hat{A} \\ &\quad + (A^a)^\top c^\top \hat{c} \hat{A}^a - (A^{b+1})^\top c^\top \hat{c} \hat{A}^{b+1} \\ &= A^\top \hat{W}_k \hat{A} + (A^a)^\top c^\top \hat{c} \hat{A}^a - (A^{b+1})^\top c^\top \hat{c} \hat{A}^{b+1} \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} \hat{\xi}_{k-1}^\top &= \sum_{i=k+a}^{k+b} \hat{c} \hat{A}^{i-k+1} y_i \\ &= \left( \sum_{i=k+a}^{k+b} \hat{c} \hat{A}^{i-k} y_i \right) \hat{A} + \hat{c} \hat{A}^a y_{k+a-1} - \hat{c} \hat{A}^{b+1} y_{k+b} \\ &= \hat{\xi}_k^\top \hat{A} + \hat{c} \hat{A}^a y_{k+a-1} - \hat{c} \hat{A}^{b+1} y_{k+b} \end{aligned} \quad (\text{A.13})$$

$$\hat{\kappa}_{k-1} = \sum_{i=k+a-1}^{k+b-1} \check{c} \check{A}^{i-k+1} y_i^2$$

$$\begin{aligned}
&= \left( \sum_{i=k+a}^{k+b} \check{c}\check{A}^{i-k} y_i^2 \right) \check{A} + \check{c}\check{A}^a y_{k+a-1}^2 - \check{c}\check{A}^{b+1} y_{k+b}^2 \\
&= \widehat{\kappa}_k \check{A} + \check{c}\check{A}^a y_{k+a-1}^2 - \check{c}\check{A}^{b+1} y_{k+b}^2
\end{aligned} \tag{A.14}$$



# Appendix B

## Derivations in Chapter 6

### B.1 Derivation of (6.7)

$$\begin{aligned}(\alpha^\top x^q)(\beta^\top x^q) &= (\alpha^\top x^q) \otimes (\beta^\top x^q) \\ &= (\alpha \otimes \beta)^\top (x^q \otimes x^q) \\ &= \tilde{\alpha}^\top x \left[ (I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q) \right]_q\end{aligned}\tag{B.1}$$

using property (6.93) of Kronecker products in the second step and (6.65) in the last step

### B.2 Derivation of (6.43)

$$\begin{aligned}\left[ \alpha^\top (x^q \otimes y^r) \right]^2 &= \text{Tr} \left( (x^q \otimes y^r)^\top \alpha \alpha^\top (x^q \otimes y^r) \right) \\ &= \text{Tr} \left( \alpha \alpha^\top (x^q \otimes y^r) (x^q \otimes y^r)^\top \right) \\ &\quad \text{using (6.123)} \\ &= \text{vec} \left( \alpha \alpha^\top \right)^\top \text{vec} \left( (x^q \otimes y^r) (x^q \otimes y^r)^\top \right) \\ &\quad \text{using (6.93)} \\ &= \text{vec} \left( \alpha \alpha^\top \right)^\top \text{vec} \left( (x^q (x^q)^\top) \otimes (y^r (y^r)^\top) \right) \\ &\quad \text{using (6.100)} \\ &= \left( \text{vec}(\alpha \alpha^\top) \right)^\top R_{Q;R} \left[ \text{vec}(x^q (x^q)^\top) \otimes \text{vec}(y^r (y^r)^\top) \right]\end{aligned}$$

$$\begin{aligned}
& \text{using (6.104)} \\
& = (\alpha \otimes \alpha)^\top R_{Q;R} [(x^q \otimes x^q) \otimes (y^r \otimes y^r)] \\
& = (R_{Q;R}(\alpha \otimes \alpha))^\top (x^{Mq} \otimes y^{M'r}) \\
& = \tilde{\alpha}^\top (x^{\tilde{q}} \otimes y^{\tilde{r}})
\end{aligned} \tag{B.2}$$

and  $R_{Q;R}$  as in (6.103), and  $M$  and  $M'$  as in (6.10).

### B.3 Derivation of (6.65)

The Kronecker product of two vector exponents writes as

$$\begin{aligned}
x^q \otimes x^r & = (x^q \otimes \mathbf{1}_R) \circ (\mathbf{1}_Q \otimes x^r) \\
& = x^{(q \otimes \mathbf{1}_R)} \circ x^{(\mathbf{1}_Q \otimes r)} \\
& = x^{(I_Q q) \otimes (\mathbf{1}_R \cdot \mathbf{1})} \circ x^{(\mathbf{1}_Q \cdot \mathbf{1}) \otimes (I_R r)} \\
& = x^{(I_Q \otimes \mathbf{1}_R)q} \circ x^{(\mathbf{1}_Q \otimes I_R)r} \\
& = x^{(I_Q \otimes \mathbf{1}_R)q + (\mathbf{1}_Q \otimes I_R)r}
\end{aligned} \tag{B.3}$$

applying (6.117) to find the first line, (6.93) to the exponents of the third line, and (6.64) to the second last line.

### B.4 Derivation of (6.37)

The sum of two vectors as in (6.37) is

$$\begin{aligned}
& \alpha^\top x^q + \beta^\top y^r \\
& = \left( \underbrace{\begin{bmatrix} 0 \\ \alpha \end{bmatrix}^\top}_{\alpha^\top x^q} x^{\tilde{q}} \otimes \underbrace{\begin{bmatrix} 1 \\ \mathbf{0}_R \end{bmatrix}^\top}_1 y^{\tilde{r}} \right) + \left( \underbrace{\begin{bmatrix} 1 \\ \mathbf{0}_Q \end{bmatrix}^\top}_1 x^{\tilde{q}} \otimes \underbrace{\begin{bmatrix} 0 \\ \beta \end{bmatrix}^\top}_{\beta^\top y^r} y^{\tilde{r}} \right) \\
& = \left( \begin{bmatrix} 0 \\ \alpha \end{bmatrix}^\top \otimes \begin{bmatrix} 1 \\ \mathbf{0}_R \end{bmatrix}^\top \right) (x^{\tilde{q}} \otimes y^{\tilde{r}}) + \left( \begin{bmatrix} 1 \\ \mathbf{0}_Q \end{bmatrix}^\top \otimes \begin{bmatrix} 0 \\ \beta \end{bmatrix}^\top \right) (x^{\tilde{q}} \otimes y^{\tilde{r}}) \\
& = \left( \left( \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathbf{0}_R \end{bmatrix} \right) + \left( \begin{bmatrix} 1 \\ \mathbf{0}_Q \end{bmatrix} \otimes \begin{bmatrix} 0 \\ \beta \end{bmatrix} \right) \right)^\top (x^{\tilde{q}} \otimes y^{\tilde{r}}) \\
& = \left( \underbrace{\left( \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathbf{0}_R \end{bmatrix} \right)}_{\tilde{\alpha}} + \underbrace{\left( \begin{bmatrix} 1 \\ \mathbf{0}_Q \end{bmatrix} \otimes \begin{bmatrix} 0 \\ \beta \end{bmatrix} \right)}_{\tilde{\beta}} \right)^\top (x^{\tilde{q}} \otimes y^{\tilde{r}})
\end{aligned} \tag{B.4}$$



with

$$\tilde{\alpha} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathbf{0}_R \end{bmatrix}, \quad \tilde{q} = \begin{bmatrix} 0 \\ q \end{bmatrix}, \quad (\text{B.5})$$

$$\tilde{\beta} = \begin{bmatrix} 1 \\ \mathbf{0}_Q \end{bmatrix} \otimes \begin{bmatrix} 0 \\ \beta \end{bmatrix}, \quad \tilde{r} = \begin{bmatrix} 0 \\ r \end{bmatrix}, \quad (\text{B.6})$$

and with  $\mathbf{0}_n = [0, \dots, 0]^\top$  of size  $n$ .

## B.5 Derivation of (6.40)

$$\begin{aligned} (\alpha^\top x^q)(\beta^\top y^r) &= (\alpha^\top x^q) \otimes (\beta^\top y^r) \\ &= (\alpha^\top \otimes \beta^\top)(x^q \otimes y^r) \\ &= (\alpha \otimes \beta)^\top(x^q \otimes y^r) \end{aligned} \quad (\text{B.7})$$

using (6.93) in the second line.

## B.6 Derivation of (6.58)

$$\begin{aligned} \alpha^\top(x^q \otimes c^s \otimes y^r) &= \alpha^\top(I_Q x^q \otimes c^s \mathbf{1} \otimes I_R y^r) \\ &= \alpha^\top(I_Q \otimes c^s \otimes I_R)(x^q \otimes \mathbf{1} \otimes y^r) \\ &= ((I_Q \otimes c^s \otimes I_R)^\top \alpha)^\top(x^q \otimes y^r) \\ &= ((I_Q \otimes (c^s)^\top \otimes I_R) \alpha)^\top(x^q \otimes y^r) \end{aligned} \quad (\text{B.8})$$

applying (6.93) in the first line.

## B.7 Derivation of (6.60)

$$\begin{aligned} \int \alpha^\top(z^s \otimes x^q \otimes y^r) dx &= \alpha^\top(z^s \otimes \left( \int x^q dx \right) \otimes y^r) \\ &= \alpha^\top(z^s \otimes (\delta \circ x^{\tilde{q}}) \otimes y^r) \\ &= \alpha^\top((\mathbf{1}^s \otimes \delta \otimes \mathbf{1}^r) \circ (z^s \otimes x^{\tilde{q}} \otimes y^r)) \\ &= (\alpha \circ (\mathbf{1}^s \otimes \delta \otimes \mathbf{1}^r))^\top(z^s \otimes x^{\tilde{q}} \otimes y^r) \end{aligned} \quad (\text{B.9})$$

using (6.120) in the last step and with  $\tilde{q}$  as in (6.19).

## B.8 Derivation of (6.71)

$$J(\lambda, y) = \int_a^b \left[ \alpha^\top x^q - \lambda \beta^\top (yx^q) \right]^2 dx \quad (\text{B.10})$$

$$= \int_a^b \left[ \alpha^\top x^q - \lambda \text{vdiag}(\beta)^\top (y^q \otimes x^q) \right]^2 dx \quad (\text{B.11})$$

$$\begin{aligned} &= \int_a^b (\alpha^\top x^q)^2 dx - 2\lambda \int_a^b \alpha^\top x^q \text{vdiag}(\beta)^\top (y^q \otimes x^q) dx \\ &\quad + \lambda^2 \int_a^b (\text{vdiag}(\beta)^\top (y^q \otimes x^q))^2 dx \\ &= \int_a^b (\alpha \otimes \alpha)^\top x^{Mq} dx - 2\lambda \int_a^b (\text{vdiag}(\beta) \otimes \alpha)^\top (y^q \otimes \underbrace{x^q \otimes x^q}_{x^{Mq}}) dx \\ &\quad + \lambda^2 \int_a^b R_{Q;Q} (\text{vdiag}(\beta) \otimes \text{vdiag}(\beta))^\top (y^{Mq} \otimes x^{Mq}) dx \quad (\text{B.12}) \end{aligned}$$

$$\begin{aligned} &= \left[ \underbrace{(\delta \circ (\alpha \otimes \alpha))^\top}_{c_1} (x^{Mq+1}) \right]_a^b \\ &\quad - 2\lambda \left[ \underbrace{((1^q \otimes \delta) \circ (\text{vdiag}(\beta) \otimes \alpha))^\top}_{c_2} (y^q \otimes x^{Mq+1}) \right]_a^b \\ &\quad + \lambda^2 \left[ \underbrace{((1^{Mq} \otimes \delta) \circ (R_{Q;Q} (\text{vdiag}(\beta) \otimes \text{vdiag}(\beta))))^\top}_{c_3} (y^{Mq} \otimes x^{Mq}) \right]_a^b \\ &= c_1^\top (b^{Mq+1} - a^{Mq+1}) - 2\lambda c_2^\top [(y^q \otimes b^{Mq+1}) - (y^q \otimes a^{Mq+1})] \\ &\quad + \lambda^2 c_3^\top [(y^{Mq} \otimes b^{Mq+1}) - (y^{Mq} \otimes a^{Mq+1})] \\ &= \tilde{\alpha}_1 - 2\lambda \tilde{\alpha}_2^\top y^q + \lambda^2 \tilde{\alpha}_3^\top y^{Mq} \quad (\text{B.14}) \end{aligned}$$

applying in line (B.11) the rule of dilation (6.56); in line (B.12) the rules (6.48), (6.42) and (6.41) with  $M = [(I_Q \otimes \mathbf{1}_Q) + (\mathbf{1}_Q \otimes I_Q)]$  as in (6.10) or (6.48); in line (B.13) the rule (6.61) with  $\delta$  according to (6.21) and  $R_{Q;Q}$  as in (6.103); and with the coefficient vectors

$$\tilde{\alpha}_1 = (b^{Mq+1} - a^{Mq+1})^\top c_1 \quad (\text{B.15})$$

$$\tilde{\alpha}_2 = (I_Q \otimes (b^{Mq+1} - a^{Mq+1})^\top) c_2 \quad (\text{B.16})$$

$$\tilde{\alpha}_3 = (I_{Q^2} \otimes (b^{Mq+1} - a^{Mq+1})^\top) c_3 . \quad (\text{B.17})$$

## B.9 Derivation of (6.82) and (6.83)

For  $m_\alpha$  we get, using (6.17),

$$\begin{aligned}
 m_\alpha &= \frac{1}{b-a} (\text{diag}(\delta)\alpha)^\top (b^{q+1} - a^{q+1}) \\
 &= \alpha^\top \underbrace{\left( \frac{1}{b-a} \text{diag}(\delta)(b^{q+1} - a^{q+1}) \right)}_{A_m} \\
 &= \alpha^\top A_m .
 \end{aligned} \tag{B.18}$$

For  $m_\beta(y)$  we get, applying in the first line (6.56),

$$\begin{aligned}
 m_\beta(y) &= \frac{1}{b-a} \int_a^b \text{vdiag}(\beta)^\top (y^q \otimes x^q) dx \\
 &= \frac{1}{b-a} (\text{diag}(1^q \otimes \delta)\beta)^\top ((y^q \otimes b^q) - (y^q \otimes a^q)) \\
 &= \frac{1}{b-a} (\text{diag}(1^q \otimes \delta)\beta)^\top (y^q \otimes (b^q - a^q)) \\
 &= \beta_m^\top y^q
 \end{aligned} \tag{B.19}$$

with  $\delta$  according to (6.21) and

$$\beta_m = \frac{1}{b-a} (I_Q \otimes (b^q - a^q)) (\text{diag}(1^q \otimes \delta)\beta) \tag{B.20}$$



# Bibliography

- [1] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, “ECG Beat Detection Using Filter Banks,” *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 192–202, 1999.
- [2] R. H. Bartels and G. W. Stewart, “Solution of the Matrix Equation  $AX + XB = C$  [F4],” *Commun. ACM*, vol. 15, no. 9, pp. 820–826, Sep. 1972.
- [3] M. N. Benzaïon, D. F. Ortega, J. M. Thierer, R. A. S. Torcivia, L. Aldunate, A. E. A. de Lima, D. Navia, A. Dorsa, A. Rossi, and M. Trivi, “Comparison of the amplitude of the p-wave from intracardiac electrocardiogram obtained by means of a central venous catheter filled with saline solution to that obtained via esophageal electrocardiogram,” *The American Journal of Cardiology*, vol. 98, no. 7, pp. 970–981, 2006.
- [4] G. Bierman, “Fixed Memory Least Squares Filtering,” *IEEE transactions on Information Theory*, vol. 21, no. 6, pp. 690–692, 1975.
- [5] C. M. Bishop, *Pattern recognition and machine learning*, ser. Information science and statistics. New York: Springer, 2006.
- [6] M. Blum, “Fixed memory least squares filters using recursion methods,” *IRE Transactions on Information Theory*, vol. 3, no. 3, pp. 178–182, Sep. 1957.
- [7] A. Bruckstein and T. Kailath, “Recursive limited memory filtering and scattering theory,” *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 440–443, May 1985.
- [8] L. Bruderer, H. Loeliger, and N. Zalmi, “Local statistical models from deterministic state space models, likelihood filtering, and local

- typicality,” in *2014 IEEE International Symposium on Information Theory*, Jun. 2014, pp. 1106–1110.
- [9] P. Buxbaum, “Fixed-memory recursive filters,” *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 113–115, 1974.
- [10] M. Cremer, “Über die direkte Ableitung der Aktionsströme des menschlichen Herzens vom Ösophagus und über das EKG des Föten,” *Munch. Med. Wochenschr.*, no. 53, pp. 811–813, 1906.
- [11] F. Daum, “Nonlinear filters: beyond the Kalman filter,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, Aug. 2005.
- [12] C. de Boor, “On calculating with B-splines,” *Journal of Approximation Theory*, vol. 6, no. 1, pp. 50–62, Jul. 1972.
- [13] P. Dierckx, *Curve and Surface Fitting with Splines*. Clarendon Press, 1995.
- [14] A. Doucet and C. Andrieu, “Iterative algorithms for state estimation of jump Markov linear systems,” *IEEE Transactions on Signal Processing*, vol. 49, no. 6, pp. 1216–1227, June 2001.
- [15] J. Durbin and S. J. Koopman, *Time series analysis by state space methods*. Oxford University Press, 2012, no. 38.
- [16] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, “Bayesian nonparametric inference of switching dynamic linear models,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1569–1585, April 2011.
- [17] K. Galkowski, “Matrix description of multivariable polynomials,” *Linear Algebra and its Applications*, vol. 234, pp. 209–226, Feb. 1996.
- [18] S. S. Haykin, *Adaptive Filter Theory*, 5th ed. Pearson, 2014.
- [19] A. Intini, R. N. Goldstein, P. Jia, C. Ramanathan, K. Ryu, B. Giannattasio, R. Gilkeson, B. S. Stambler, P. Brugada, W. G. Stevenson, Y. Rudy, and A. L. Waldo, “Electrocardiographic imaging (ECGI), a novel diagnostic modality used for mapping of focal left ventricular tachycardia in a young athlete,” *Heart Rhythm*, vol. 2, no. 11, pp. 1250–1252, 2005.

- [20] A. Jazwinski, "Limited memory optimal filtering," *IEEE Transactions on Automatic Control*, vol. 13, no. 5, pp. 558–563, Oct. 1968.
- [21] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
- [22] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.
- [23] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of Fluids Engineering*, vol. 83, no. 1, pp. 95–108, 1961.
- [24] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [25] B. U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42–57, Jan. 2002.
- [26] D. Lee, M. Morf, and B. Friedlander, "Recursive least squares ladder estimation algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 627–641, Jun 1981.
- [27] G. Lenis, N. Pilia, A. Loewe, W. H. W. Schulze, and O. Dössel, "Comparison of Baseline Wander Removal Techniques considering the Preservation of ST Changes in the Ischemic ECG: A Simulation Study," 2017.
- [28] C. F. V. Loan, "The ubiquitous Kronecker product," *Journal of Computational and Applied Mathematics*, vol. 123, no. 1, pp. 85–100, Nov. 2000.
- [29] H. Loeliger, L. Bolliger, C. Reller, and S. Korl, "Localizing, forgetting, and likelihood filtering in state-space models," in *2009 Information Theory and Applications Workshop*, Feb. 2009, pp. 184–186.
- [30] H. A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmi, "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing," in *Information Theory and Applications Workshop (ITA)*, Jan. 2016, pp. 1–10.

- [31] A. Logothetis and V. Krishnamurthy, "Expectation maximization algorithms for MAP estimation of jump Markov linear systems," *IEEE Transactions on Signal Processing*, vol. 47, no. 8, pp. 2139–2156, Aug 1999.
- [32] P. W. Macfarlane, A. Van Oosterom, O. Pahlm, P. Kligfield, M. Janse, and J. Camm, *Comprehensive electrocardiology*. Springer Science & Business Media, 2010.
- [33] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.
- [34] B. J. Messinger-Rapport and Y. Rudy, "Noninvasive recovery of epicardial potentials in a realistic heart-torso geometry. Normal sinus rhythm." *Circulation research*, vol. 66, no. 4, pp. 1023–1039, 1990.
- [35] S. A. Mortier, "Esophageal isopotential maps, a case series," MD dissertation, University of Berne, Medical Faculty, 2018.
- [36] T. Niederhauser, T. Marisa, A. Haeberlin, J. Goette, M. Jacoment, and R. Vogel, "High-resolution esophageal long-term ECG allows detailed atrial wave morphology analysis in case of atrial ectopic beats," *Med Biol Eng Comput*, vol. 50, no. 7, pp. 769–772, Jul 2012.
- [37] D. Pollock, "On Kronecker products, tensor products and matrix differential calculus," *International Journal of Computer Mathematics*, vol. 90, no. 11, pp. 2462–2476, Nov. 2013.
- [38] C. Reller, H. Loeliger, and S. Maranò, "Multi-sensor estimation and detection of phase-locked sinusoids," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 3872–3875.
- [39] C. Reller, "State-space methods in statistical signal processing: New ideas and applications," Ph.D. dissertation, No. 20584, ETH Zurich, 2012.
- [40] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [41] R. W. Schafer, "What is a Savitzky–Golay filter?" *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111–117, July 2011.



- [42] I. J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions. Part B: on the problem of osculatory interpolation. a second class of analytic approximation formulae," *Quarterly of Applied Mathematics*, vol. 4, no. 2, pp. 112–141, 1946.
- [43] G. Turin, "An introduction to matched filters," *IRE Transactions on Information Theory*, vol. 6, no. 3, pp. 311–329, June 1960.
- [44] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, Nov. 1999.
- [45] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing*. Cambridge: Cambridge University Press, 2014.
- [46] R. A. Wildhaber, "System and method for providing a representation of cardiac cell charge state," 7 2015, patent CF/00961/15.
- [47] R. A. Wildhaber, N. Zalmai, M. Jacomet, and H. Loeliger, "Windowed State-Space Filters for Signal Detection and Separation," *IEEE Transactions on Signal Processing*, vol. 66, no. 14, pp. 3768–3783, Jul. 2018.
- [48] R. A. Wildhaber, N. Zalmai, M. Jacomet, and H. A. Loeliger, "Signal Detection and Discrimination for Medical Devices Using Windowed State Space Filters," in *13th IASTED International Conference on Biomedical Engineering (BioMed)*, Feb 2017, pp. 125–133.
- [49] R. Wildhaber, S. Mortier, A. Haerberlin, R. Sweda, D. Bruegger, T. Niederhauser, M. Jacomet, J. Goette, and H. Tanner, "A novel 2-D spatial-temporal ECG representation using multipolar esophageal catheters: A pilot study," *JACC: Clinical Electrophysiology*, vol. 3, no. 10, Supplement, p. S2, 2017, Venice Arrhythmias 2017 Abstracts.
- [50] R. A. Wildhaber, D. Bruegger, N. Zalmai, H. Malmberg, J. Goette, M. Jacomet, H. Tanner, A. Haerberlin, and H.-A. Loeliger, "Estimation of the cardiac field in the esophagus using a multipolar esophageal catheter," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 4, pp. 791–800, 2018.

- 
- [51] S. B. Wilson and R. Emerson, “Spike detection: a review and comparison of algorithms,” *Clinical Neurophysiology*, vol. 113, no. 12, pp. 1873–1881, Dec. 2002.
- [52] L. A. Zadeh and C. A. Desoer, *Linear System Theory: The State Space Approach*. R. E. Krieger Publishing Company, 1963.
- [53] N. Zalmi, C. Kaeslin, L. Bruderer, S. Neff, and H.-A. Loeliger, “Gesture recognition from magnetic field measurements using a bank of linear state space models and local likelihood filtering,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.
- [54] N. Zalmi, R. A. Wildhaber, D. Clausen, and H.-A. Loeliger, “Inferring depolarization of cells from 3D-electrode measurements using a bank of linear state space models,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.
- [55] N. Zalmi, R. A. Wildhaber, and H.-A. Loeliger, “Autonomous state space models for recursive signal estimation beyond least squares,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 341–345.
- [56] N. Zalmi, “A State Space World for Detecting and Estimating Events and Learning Sparse Signal Decompositions,” Ph.D. dissertation, No. 24360, ETH Zurich, 2017.

# About the Author

Reto Andreas Wildhaber was born in Baar, Switzerland in 1979. He attended high school in Zug, Switzerland, and did an apprenticeship as an electronics technician. In 2003, he received the degree in electrical engineering from the University of Applied Sciences, Rapperswil, Switzerland, and in 2013 the M.Sc. degree in medicine and in 2014 the M.D. degree, both from the University of Zurich, Switzerland. Since 2014, he has been a PhD candidate at the Signal and Information Processing Laboratory (ISI) at ETH Zurich and is employed at Bern University of Applied Sciences, Biel. For clinical research, he collaborates with the University Hospital Bern. Since 2019, he holds an assistant professor position (tenure track) at Bern University of Applied Sciences.



## Series in Signal and Information Processing

edited by Hans-Andrea Loeliger

- Vol. 1: Hanspeter Schmid, **Single-Amplifier Biquadratic MOSFET-C Filters**. ISBN 3-89649-616-6
- Vol. 2: Felix Lustenberger, **On the Design of Analog VLSI Iterative Decoders**. ISBN 3-89649-622-0
- Vol. 3: Peter Theodor Wellig, **Zerlegung von Langzeit-Elektromyogrammen zur Prävention von arbeitsbedingten Muskelschäden**. ISBN 3-89649-623-9
- Vol. 4: Thomas P. von Hoff, **On the Convergence of Blind Source Separation and Deconvolution**. ISBN 3-89649-624-7
- Vol. 5: Markus Erne, **Signal Adaptive Audio Coding using Wavelets and Rate Optimization**. ISBN 3-89649-625-5
- Vol. 6: Marcel Joho, **A Systematic Approach to Adaptive Algorithms for Multichannel System Identification, Inverse Modeling, and Blind Identification**. ISBN 3-89649-632-8
- Vol. 7: Heinz Mathis, **Nonlinear Functions for Blind Separation and Equalization**. ISBN 3-89649-728-6
- Vol. 8: Daniel Lippuner, **Model-Based Step-Size Control for Adaptive Filters**. ISBN 3-89649-755-3
- Vol. 9: Ralf Kretzschmar, **A Survey of Neural Network Classifiers for Local Wind Prediction**. ISBN 3-89649-798-7
- Vol. 10: Dieter M. Arnold, **Computing Information Rates of Finite State Models with Application to Magnetic Recording**. ISBN 3-89649-852-5
- Vol. 11: Pascal O. Vontobel, **Algebraic Coding for Iterative Decoding**. ISBN 3-89649-865-7
- Vol. 12: Qun Gao, **Fingerprint Verification using Cellular Neural Networks**. ISBN 3-89649-894-0
- Vol. 13: Patrick P. Merkli, **Message-Passing Algorithms and Analog Electronic Circuits**. ISBN 3-89649-987-4
- Vol. 14: Markus Hofbauer, **Optimal Linear Separation and Deconvolution of Acoustical Convolutional Mixtures**. ISBN 3-89649-996-3
- Vol. 15: Sascha Korl, **A Factor Graph Approach to Signal Modelling, System Identification and Filtering**. ISBN 3-86628-032-7
- Vol. 16: Matthias Frey, **On Analog Decoders and Digitally Corrected Converters**. ISBN 3-86628-074-2
- Vol. 17: Justin Dauwels, **On Graphical Models for Communications and Machine Learning: Algorithms, Bounds, and Analog Implementation**. ISBN 3-86628-080-7

- Vol. 18: Volker Maximilian Koch, **A Factor Graph Approach to Model-Based Signal Separation.** ISBN 3-86628-140-4
- Vol. 19: Junli Hu, **On Gaussian Approximations in Message Passing Algorithms with Application to Equalization.** ISBN 3-86628-212-5
- Vol. 20: Maja Ostojic, **Multitree Search Decoding of Linear Codes.** ISBN 3-86628-363-6
- Vol. 21: Murti V.R.S. Devarakonda, **Joint Matched Filtering, Decoding, and Timing Synchronization.** ISBN 3-86628-417-9
- Vol. 22: Lukas Bolliger, **Digital Estimation of Continuous-Time Signals Using Factor Graphs.** ISBN 3-86628-432-2
- Vol. 23: Christoph Reller, **State-Space Methods in Statistical Signal Processing: New Ideas and Applications.** ISBN 3-86628-447-0
- Vol. 24: Jonas Biveroni, **On A/D Converters with Low-Precision Analog Circuits and Digital Post-Correction.** ISBN 3-86628-452-7
- Vol. 25: Georg Wilckens, **A New Perspective on Analog-to-Digital Conversion of Continuous-Time Signals.** ISBN 3-86628-469-1
- Vol. 26: Jiun-Hung Yu, **A Partial-Inverse Approach to Decoding Reed-Solomon Codes and Polynomial Remainder Codes.** ISBN 3-86628-527-2
- Vol. 27: Lukas Bruderer, **Input Estimation and Dynamical System Identification: New Algorithms and Results.** ISBN 3-86628-533-7
- Vol. 28: Sarah Neff, **A New Approach to Information Processing with Filters and Pulses.** ISBN 3-86628-575-2
- Vol. 29: Christian Schürch, **On Successive Cancellation Decoding of Polar Codes and Related Codes.** ISBN 3-86628-580-9
- Vol. 30: Nour Zalmi, **A State Space World for Detecting and Estimating Events and Learning Sparse Signal Decompositions.** ISBN 978-3-86628-594-1
- Vol. 31: Federico Wadehn, **State Space Methods with Applications in Biomedical Signal Processing.** ISBN 978-3-86628-640-5