Title
Convex Optimization Methods for System Identification with Applications to Noninvasive Intracranial Pressure Estimation

Author
Gunn, Cameron Allan

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Convex Optimization Methods for System Identification with Applications to Noninvasive Intracranial Pressure Estimation

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical and Computer Engineering

by

Cameron Allan Gunn

2018

ABSTRACT OF THE DISSERTATION

Convex Optimization Methods for System Identification with
Applications to Noninvasive Intracranial Pressure Estimation

by

Cameron Allan Gunn
Doctor of Philosophy in Electrical and Computer Engineering
University of California, Los Angeles, 2018
Professor Lieven Vandenberghe, Chair

After a traumatic brain injury, it is important for some patients' *intracranial pressure* (ICP) to be measured while they are in intensive care. However, monitoring ICP first requires an invasive surgical procedure, an impediment that has prompted research on noninvasive ICP (NICP) estimation. This dissertation examines NICP estimation from the perspective of linear dynamical systems, and presents methods to address some of the challenges that have limited the success of NICP estimators. Examples of these challenges include the unreliability of corrupted signal data, and a large inter-patient variability that limits the ability to compare new patients to past patients.

Three sets of methods are presented in this dissertation. The first methods mitigate the effect of corruptions in *cerebral blood flow velocity* signals, which are strong predictors of ICP, but often contain artifacts or sections of missing data. These methods find completed approximations of these signals that produce low-order systems. The second family of methods provide an approach for *clustering* linear dynamical systems by their behavior. This framework can be used to determine a subset of past patients with similar signal dynamics to a new patient. The final methods are a novel approach to NICP estimation in which the partially-available data of a new patient is combined with information from past patients whose ICP was invasively measured.

The methods presented are flexible in two respects. First, they are developed for general

linear dynamical systems, and so can be adapted to any new applications where these models are used. Second, they are posed as convex optimization problems, which can be easily extended to new scenarios through the use of new weights, constraints, and penalty functions. The methods are solved using proximal algorithms, a family of first-order convex optimization algorithms, which result in computationally tractable formulations.

The dissertation of Cameron Allan Gunn is approved.

Vwani P Roychowdhury

Paul M Vespa

Xiao Hu

Lieven Vandenberghe, Committee Chair

University of California, Los Angeles

2018

For Teni

TABLE OF CONTENTS

1 Introduction **1**

1.1 Framing the NICP estimation problem 2

1.2 Contributions of this dissertation 4

1.3 Outline of this dissertation 5

2 Review of noninvasive ICP estimation **7**

2.1 ICP monitoring in critical care 7

2.1.1 Clinical need for ICP monitoring 7

2.1.2 Intraventricular catheterization 8

2.1.3 Invasive alternatives to IVC 10

2.1.4 Clinical analysis of ICP 10

2.2 Noninvasive ICP estimation 11

2.2.1 TCD ultrasonography 11

2.2.2 Model-based NICP estimators 13

3 Review of subspace system identification **15**

3.1 Linear time-invariant systems 15

3.2 Subspace system identification 16

3.2.1 Overview of subspace identification process 17

3.2.2 Data matrices that reveal system order 19

3.3 Trace norm minimization 21

3.3.1 Applications of trace norm minimization 22

4 Enhancing corrupted physiological signals by trace norm regularization **24**

4.1 Introduction 24
4.1.1 Motivation 24
4.1.2 Background on current methods 25
4.2 Model validation 27
4.2.1 ICP data 28
4.2.2 Model configurations 28
4.2.3 Simulation and results 29
4.3 Methods for artifact rejection and missing data imputation 31
4.3.1 Basic method 31
4.3.2 Structured artifact rejection 32
4.3.3 Data-dissimilarity penalties 34
4.3.4 Smoothing penalties 35
4.3.5 Exploiting prior information 35
4.3.6 System identification 36
4.3.7 Algorithms 37
4.4 Simulations 38
4.4.1 Missing data 38
4.4.2 Artifact rejection 38
4.4.3 Structured artifact rejection 41
4.5 Discussion 49

5 Methods to cluster patient models as dynamical systems 50

5.1 Introduction 50
5.2 Background 52
5.2.1 Data model 52

5.2.2 Metrics to compare dynamical systems . . . 52
5.2.3 Background on system clustering and classification . . . 54
5.2.4 Outline of new approach . . . 54
5.3 Method development . . . 54
5.3.1 Review: Clustering vectors by subspace . . . 54
5.3.2 Clustering matrices by column space . . . 57
5.3.3 LTI model clustering method . . . 58
5.3.4 Algorithm . . . 59
5.4 Simulations . . . 59
5.4.1 Clustering validation . . . 59
5.4.2 NICP estimation . . . 60
5.5 Extensions . . . 63
5.5.1 Effect of λ . . . 63
5.5.2 Choice of group sparsity penalty . . . 66
5.5.3 Data matrix methods . . . 68
5.6 Comparison to distance matrix–based methods . . . 69
5.7 Discussion . . . 70

6 Dictionary-based signal estimation with application to NICP estimation . . . 73

6.1 Introduction . . . 73
6.1.1 Background on current methods . . . 74
6.2 Dictionary-based signal estimation . . . 75
6.2.1 The DBSE framework . . . 75
6.2.2 Choosing simulation weights . . . 79

6.2.3 Extensions . . . 80
6.2.4 Algorithm . . . 83
6.3 Simulations . . . 83
6.4 Discussion . . . 84

7 Algorithms . . . 86
7.1 Introduction . . . 86
7.2 Review of proximal operators . . . 87
7.2.1 Definition and interpretation . . . 87
7.2.2 Useful properties of proximal operators . . . 88
7.2.3 Examples of proximal operators . . . 89
7.3 Review of Douglas–Rachford splitting algorithms . . . 91
7.3.1 Alternating direction method of multipliers . . . 91
7.3.2 Primal–dual Douglas–Rachford algorithm . . . 93
7.3.3 Chambolle–Pock algorithm . . . 93
7.4 Artifact rejection and missing data imputation . . . 94
7.4.1 Basic method . . . 94
7.4.2 Structured artifact rejection . . . 96
7.5 System clustering . . . 98
7.5.1 Problem rearrangement . . . 98
7.5.2 Proximal operators . . . 100
7.6 Dictionary-based signal estimation . . . 100

8 Conclusions . . . 102

References . . . 104

LIST OF FIGURES

1.1 An overview of the NICP estimation problem. A patient's ABP, CBFV, and ICP signals can be treated as a linear dynamical system. However, both the output (ICP) signal and the system itself are missing. A dictionary of models $\Sigma_1, \ldots, \Sigma_L$ from past patients with complete information is used to help estimate the missing ICP signal. 3

2.1 Anatomy of cerebral spinal fluid (CSF) around the brain, and in the ventricular space. Source: *Blausen.com staff (2014). "Medical Gallery of Blausen Medical 2014". WikiJournal of Medicine.* 8

2.2 A CBFV waveform measured by transcranial Doppler ultrasonography. 12

2.3 A popular input–output configuration for model-based NICP estimators. 13

4.1 Examples of CBFV waveforms with various types of large artifacts. 26

4.2 Input–output configurations of the CBFV and ICP models. 29

4.3 Simulated output of the ICP model and the CBFV model in two separate episodes. 30

4.4 Imputation of missing CBFV using the CBFV model. Data was removed between 3–5 s. These results can be achieved without knowing patient ICP. 39

4.5 Imputation of missing CBFV using the ICP model. Data was removed between 3–5 s. These results require knowledge of patient ICP. 40

4.6 Median absolute imputation error by the CBFV and ICP models for different lengths of missing data. 41

4.7 Removal of a large artifact using the CBFV model. Only ABP and CBFV are required to achieve this quality of rejection. 42

4.8 Removal of a sparse, sharp artifact using the CBFV model. 43

4.9 Removal of frequent disturbances using the CBFV model. 44

4.10 Artifact rejection for varying values of λ, using the CBFV model. Lower values of λ produce signals with lower-order dynamics, while higher values more closely resemble the measured data. 45

4.11 Outcomes with prior knowledge that an artifact occurs in the range $t \in (4.5, 6.5)$ s, using the CBFV model. The clean data is kept exact, and the corrupted section is optimized over. 46

4.12 Overview of CBFV structured artifact removal via (4.12), using $q = 20$ and $\lambda = 0.35$. 47

4.13 Structure of nonzero residual *atoms* produced by the CBFV structural artifact rejection method (4.12). These six nonzero residuals start at $t = 0.25, 1.43, 2.28, 3.13, 3.85, 4.68$ s and are 0.5 s in length. 48

5.1 Edge weights between each model and corresponding cluster assignments. Models $1, \ldots, 10$ correspond to Patient 1, $11, \ldots, 20$ to Patient 2, etc.. 92% of the models were correctly clustered. 61

5.2 Adjacency of episodes in V via the Fruchterman–Reingold force–directed algorithm [FR91]. Nodes are colored by patient ID. The incorrectly-clustered nodes are marked by red crosses. Some lower-weight edges are omitted for clarity. This diagram was constructed using the NetworkX package [HSS08]. 62

5.3 The relationship between mean CBFV and mean ICP is in general poorly correlated. However, the intra-cluster correlations reveal moderately improved correlations. This result suggests that applying system clustering as a preprocessing step to NICP estimation can improve outcomes. 64

5.4 Effect of λ on the sparsity pattern of V for $L = 60$ models drawn from 6 patients. 65

5.5 The accuracy of system clustering for three different choices of group sparsity norm penalty, and models of order $n = 1, \ldots, 6$. The $L = 100$ models were constructed from 10 different patients for this experiment. While the trace norm underperforms its alternatives, there is otherwise no strong trends in the choices. Low-order models appear to work acceptably, and require substantially less computation than higher-order models. 67

5.6 Distance matrices D_M and D_R, and the corresponding clusters that were assigned via agglomerative hierarchical clustering. Neither of the cluster assignments is able to recover the underlying per-patient clusters, severely underperforming the method of (5.19) based on sparse column space clustering. 71

6.1 An overview of the dictionary-based signal estimation problem setup. Input signals $\hat{u}_{\text{meas}}(k) \in \mathbb{R}^m$ are measured, but the system $\hat{\Sigma}$ and the output $\hat{y}(k) \in \mathbb{R}$ are not. A dictionary of L similar systems $\Sigma_1, \ldots, \Sigma_L$ with complete information is available. Estimates of these systems' outputs can be generated by subjecting them to the known input $\hat{u}_{\text{meas}}(k)$; these simulated outputs are each denoted $y_{i,\text{sim}}(k)$. 76

6.2 Simulation of an ICP signal, assuming zero initial conditions. The transient region $t \in (0, 3)$ s is identified and excluded when solving (6.7). 81

6.3 Example of an NICP estimate using the DBSE method. 84

7.1 Shape of the soft thresholding function, which is the element-wise proximal operator for the ℓ_1-norm. 90

LIST OF TABLES

4.1 Simulation error the CBFV and ICP models, for both patient-specific and population-wide modeling cases. Errors are given in median absolute error, expressed as a percentage. 30

5.1 NICP estimation error from two linear predictors. The first is trained and tested on independent patients in the population. The second is trained and tested on independent patients in each of $k = 3$ clusters. The latter improved performance. 63

6.1 Median [IQR] absolute error of ICP estimates of the DBSE method in the testing set, for different weighting schemes. 84

ACKNOWLEDGMENTS

I would like to first thank my advisor, Lieven Vandenberghe, whose advice and direction throughout my PhD has been incredible. Thank you for all of the time you have graciously spent mentoring me.

Thank you to Xiao Hu, for all of your assistance with my research. Thank you to my committee members, Paul Vespa and Vwani Roychowdhury, for your time and helpful feedback. Thank you to my lab: Yifan Sun, Daniel O'Connor, Suzi Chao, Jinchao Li, and Xin Jiang, for being great people to discuss ideas with. Thank you to Geoff Chase, Richard Brown, and Mike Plank, whose advice and support are what got me here in the first place. Thank you to Fulbright New Zealand for your financial support. Thank you to Deeona Columbia, Ryo Arreola, and Mandy Smith for your help over these past five years. Thank you to Brendan Maione-Downing and Peter Fontana for your mentorship at Blue State Digital.

Thank you to my family. To Abayomi, Olufunke, and Konyinsope Adewumi: Thank you for your love and support during my studies. To my parents, Phil and Barbara Gunn: Thank you for *everything*. Thank you to Rory Gunn for being a fantastic brother.

And, of course: Thank you to my wife and best friend, Teniope Adewumi-Gunn. Your endless love, support, patience, and advice means the world to me.

VITA

2013	B.E.(Hons.), First Class Honours (Mechatronics Engineering), Te Whare Wānanga o Waitaha, Ōtautahi, Aotearoa / University of Canterbury, Christchurch, New Zealand
2015	M.S. (Electrical Engineering), UCLA, Los Angeles, California
2016	Data Science Fellow, Blue State Digital, New York, New York
2017	Quantitative Trader Intern, Sun Trading LLC, New York, New York

CHAPTER 1

Introduction

Traumatic brain injury (TBI) is the leading cause of death in young people in the United States [FGC12]. Defined as a blow, bump, jolt, or penetration of the head that disrupts normal brain function, TBI accounts for 1.7 million injuries, 275,000 hospitalizations, and 52,000 deaths in the US annually [RLT06]. Studies have estimated the annual costs associated with TBI to be $64 billion in the US [FGC12] and $40 billion in Europe [OGS12].

The total effect of neurological damage from TBI can continue for hours or days after the initial injury [BMW77, HA12]. Sources for this ongoing damage include swelling or bleeding in the brain, infection, and seizures [CMK93, USK04]. So, the critical care of TBI patients is important to their recovery; studies have shown that targeted care protocols can reduce mortality and improve patient outcomes [FTW04, AHT10].

One important quantity to measure during TBI care is *intracranial pressure* (ICP), the fluid pressure inside the skull [CP04]. Elevated ICP, or *intracranial hypertension*, can indicate swelling, lesions, or worsened cerebral perfusion of the resources necessary for recovery, and can prompt decisive clinical action to correct [SM14]. However, the process of measuring ICP is particularly invasive, and involves the insertion of a catheter through the skull and brain tissue, carrying a level of clinical risk [AKG86]. Physicians are therefore given a tradeoff between performing this invasive procedure or not knowing a patient's ICP. This tradeoff has given rise to a new, ongoing field of research: noninvasive ICP (NICP) estimation.

NICP estimation has proven to be a difficult problem, and no solutions exist that have met a clinical standard of accuracy [RBA12]. There are many approaches that have given reasonable estimates, however, from clinical heuristics to computer-driven models [RAS03, TA07, KVN12].

This dissertation will examine the NICP estimation problem from the perspective of *linear dynamical systems* theory. New methods will be developed around this framework to address some of the challenges that have limited the success of NICP estimators. Examples of these issues include the unreliability of corrupted signal data, and a large inter-patient variability that limits the opportunity to compare new patients to past patients.

1.1 Framing the NICP estimation problem

The scope of the NICP problem, as it is considered in this work, is as follows. A patient's ICP is an unknown, time-varying signal. The aim of the task is to estimate this signal. All that is measured of a patient are two 'noninvasive' time-varying signals: their *cerebal blood flow velocity* (CBFV), and their *arterial blood pressure* (ABP). It is assumed that the relationship between ABP, CBFV, and ICP can be accurately modeled by a low-order (that is, low-complexity) linear dynamical system, in which ABP and CBFV are the inputs, and ICP is the output. However, this system is unknown. Because only the system inputs are measured, and the system and its output are both unknown, this is a *blind signal estimation* problem that requires additional information or assumptions to solve.

Also available is a *dictionary* of past patient information. These patients had their ICP measured by invasive means, and so a complete picture is available: their input data, output (ICP) data, and system model are all known. So, the NICP problem is to combine (a) the incomplete information on a patient and (b) the complete information of other patients. Figure 1.1 provides a simple diagram of this framework.

There are also additional practical challenges that are specific to this data set. First, the known CBFV signals are not reliable, as they are often corrupted by artifacts or missing data due to patient motion and clinical disturbances. Second, there is large inter-patient variability across the population, and so 'one size fits all' population-wide models do not provide accurate ICP estimates. Because of these challenges, this task is a hard, unsolved problem.

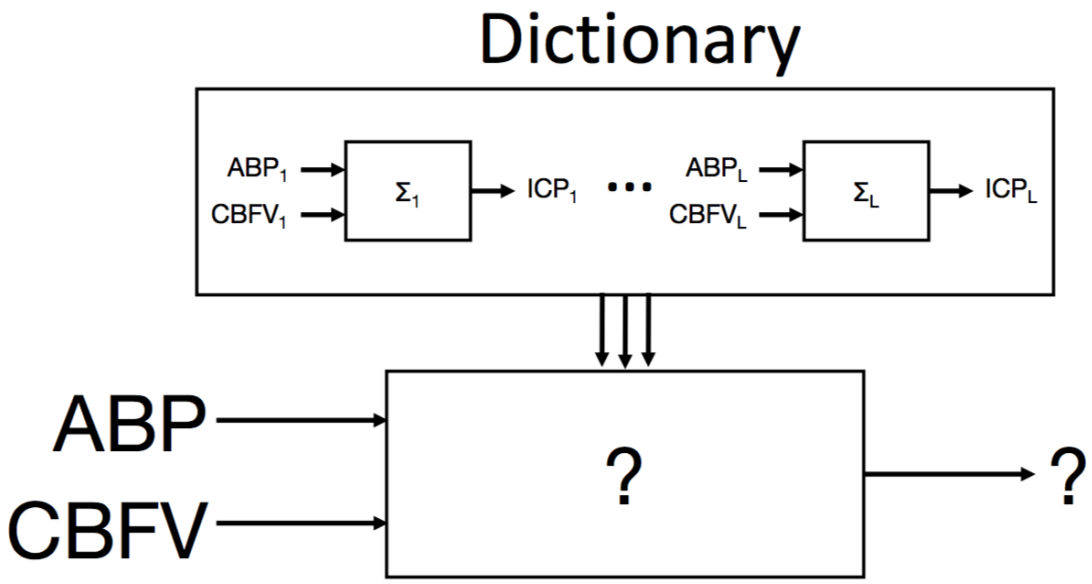

Figure 1.1: An overview of the NICP estimation problem. A patient's ABP, CBFV, and ICP signals can be treated as a linear dynamical system. However, both the output (ICP) signal and the system itself are missing. A dictionary of models $\Sigma_1, \ldots, \Sigma_L$ from past patients with complete information is used to help estimate the missing ICP signal.

1.2 Contributions of this dissertation

The primary aim of this dissertation is to develop a framework of data methods that add value to NICP estimators and the associated field of research. The secondary aim is to frame these data methods as more general problems that can be applied to many different scenarios and fields of work, maximizing their usefulness. Through these aims, the main contributions of the dissertation are in three suites of methods for operating on time series data and linear dynamical models, and their application to both the NICP problem and ICP data.

The first set of methods are designed to mitigate the effect of corruptions in biomedical signal data. These corruptions, such as large artifacts and missing data, are common in *cerebral blood flow velocity* (CBFV) signals, which are frequently used in NICP estimators. The new methods implicitly model available biomedical signals with a linear dynamical system. Approximations of these signals are then found that are complete and have dynamics of low complexity.

The second group of methods are designed to overcome the substantial inter-patient variability observed in this population. This variability has substantially limited the usefulness of population-wide predictive models, increasing the complexity of the NICP estimation task. The new methods are able to *cluster* patient models by the similarity of their dynamics. By using these methods, more-targeted models of new patients can be estimated by comparing their behavior to just a subset of similar past patients.

The final set of methods are a simple, new approach to estimating ICP noninvasively. They operate by considering both the partial information known about the current patient—that is, what can be measured noninvaively—alongside the complete information about a dictionary of past patients whose ICP was invasively measured.

All of the methods presented in the following chapters are flexible in two important respects. First, they are expressed in the language of linear dynamical systems, and can be readily adapted to other applications, from engineering to the social sciences, that use these models for time series data. Second, the methods are posed as convex optimization problems. It is often easy to extend these convex formulations to new scenarios through new

weights, penalties, and constraints. Furthermore, the convexity of these problems leads to tractable solutions.

1.3 Outline of this dissertation

The next two chapters will review background material on both the ICP problem and the technical tools that will be used. Specifically:

- **Chapter 2** will review ICP monitoring from a clinical standpoint. The need for ICP monitoring, and the associated risks, will be discussed. A range of available ICP monitoring techniques are reviewed, including traditional invasive methods, noninvasive clinical heuristics, and recent work on model-based noninvasive estimators.
- **Chapter 3** will review some underlying technical topics that are used throughout this dissertation. This includes a review of linear dynamical systems, and the family of subspace system identification methods that can be used to identify them from data using the subspaces of matrices.

The three chapters that follow will present the main contributions of this dissertation:

- **Chapter 4** presents a suite of methods to remove artifacts and impute sections of missing data in biomedical signals. The methods exploit low-order dynamical relationships between signals, fitting 'simple' models to the data using convex regularization terms that are robust to large artifacts and missing data. The methods are applied to CBFV signals, which are frequently used in NICP estimators, but are susceptible to such corruptions.
- **Chapter 5** presents a methodology for clustering dynamical systems. The methods compare systems by the columns spaces spanned by their observability matrices. This approach is tested on models of ICP dynamics and simple NICP estimators.
- **Chapter 6** proposes an approach for estimating ICP called *dictionary-based signal estimation*, that combines patient data with that of a dictionary of past patients.

These methods are applied to NICP estimation as well as other benchmark time series data.

The final body chapter provides implementation details that apply across the dissertation:

- **Chapter 7** reviews a class of convex optimization algorithms called proximal algorithms, and demonstrates how they can be used to realize the methods in Chapters 4–6.

CHAPTER 2

Review of noninvasive ICP estimation

The key motivation for this dissertation is to improve *noninvasive intracranial pressure* estimation in the critical care setting, particularly for TBI patients. This chapter will provide the clinical background to this problem: why intracranial pressure is important to measure in this population, the drawbacks of its invasive measurement process, and the progress of noninvasive alternatives that are of current research interest.

2.1 ICP monitoring in critical care

2.1.1 Clinical need for ICP monitoring

TBI results in neurological damage. Significant damage does not only occur *during* the trauma but also over several days following the event, damage that is referred to as *secondary effects*. These effects can be mitigated or prevented with targeted care protocols [HA12, BMW77]. The Brain Trauma Foundation (BTF) [CTO17] guidelines for TBI care suggest that three values should be monitored: arterial blood pressure (ABP), intracranial pressure (ICP), and cerebral perfusion pressure (CPP). ABP is the pressure exerted throughout the body's arterial system. ICP is the pressure exerted on the fluids in the skull, which comprise around 5% of total skull volume, as shown in Fig. 2.1. CPP is the pressure gradient of blood flow between the arterial and cerebral systems that drives blood delivery to the brain [Neu00].

ABP measurement is routine in clinical care and so its monitoring does not introduce an additional clinical decision [LST13]. ICP monitoring however is not routine, and requires specific indication before it is performed [NKB82]. CPP is a function of ABP and ICP, and

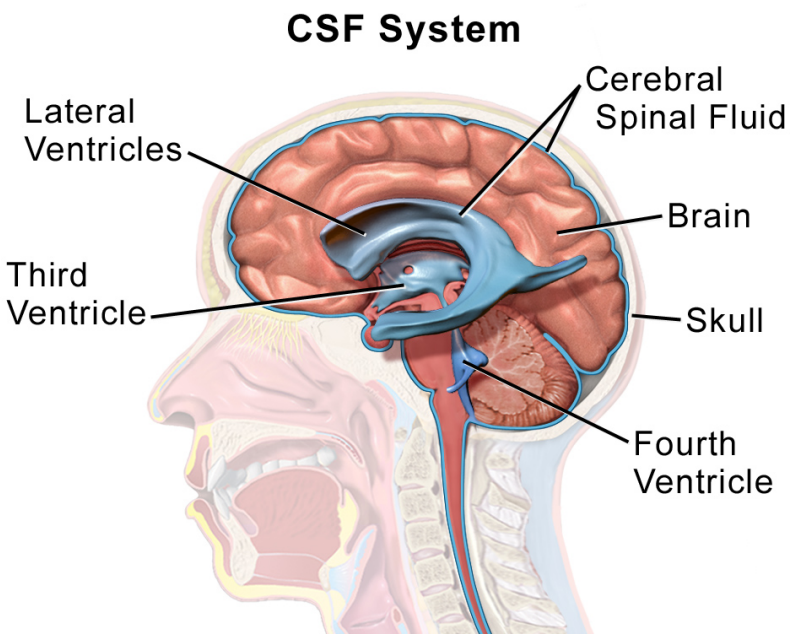

Figure 2.1: Anatomy of cerebral spinal fluid (CSF) around the brain, and in the ventricular space. Source: *Blausen.com staff (2014). "Medical Gallery of Blausen Medical 2014". WikiJournal of Medicine.*

can be calculated when both measurements are present [Neu00].

ICP is in itself a useful measurement, for understanding swelling and detecting intracranial lesions. However, it is primarily measured because it facilitates therapies to improve CPP and maintain adequate supply of resources to the brain during recovery [HA12]. Low cerebral perfusion can be treated through a variety of methods: draining the cerebrospinal fluid (CSF) to reduce ICP, administering vasopressors to raise ABP, and where necessary, giving mannitol to further reduce ICP [RRJ95]. Hence, ICP monitoring can lead to actionable steps during care.

2.1.2 Intraventricular catheterization

Intraventricular catheterization (IVC) is the most widely-used method of ICP monitoring [DJS12]. It is accurate, and is used as the gold standard to benchmark all other measurement devices and methods [ZDP13].

IVC is performed by inserting a catheter through the skull and brain tissue, into the *ventricular system* of cavities in the brain, which are responsible for production of CSF and autoregulation of blood flow to the brain. The gradient between intracranial pressure and atmospheric pressure causes the CSF to flow through the catheter, which is then measured by a calibrated strain gauge to determine ICP. An IVC can also be used for CSF management by draining excess fluid, meaning that high ICP readings are immediately actionable. IVC has been widely adopted because of its high accuracy, low cost, and ability to drain or test CSF [CTO17].

Despite the benefits of using IVC, the deep penetration of the brain carries risk to the patient, dissuading its universal practice. The incidence of this risk is unclear: Park et al. surveyed reports that claim between 0–45% infection rates attributed to the catheterization process [PGK04]. While there is a body of literature reporting increased infection from ICP monitoring [MAL84, AKG86, KWP85], recent studies have found more limited evidence [WRK93]. Other reported risks include hemorrhage, meningitis, and complications from misuse, malfunction, or malpositioning [GDD98, PT94, SNS88, NKB82, MWS98, Neu11]. The risks associated with ICP monitoring are reported to be heightened in patients with liver failure [KSB12], fulminant hepatic failure [ABK08, LSK00], or pregnant women with preeclampsia [SBH02]. Additionally, IVC is not guaranteed to succeed, and can fail due to blood clots, air bubbles, and other debris in the skull that were caused by the TBI [ZDP13].

Because of these risks, the BTF recommends ICP monitoring only when patients are at risk of elevated ICP, the data is useful, and monitoring and treatment would improve outcomes [CTO17]. Specifically, the BTF guidelines indicate ICP monitoring for TBI patients with an abnormal CT scan (revealing contusions or swelling, for example) and a *Glasgow Coma Scale* (GCS) score of 3–8, which is taken from a possible range of 3 (deep comatose) to 15 (fully awake) [TJ74]. Clinical options are also suggested for a subset of similar cases.

2.1.3 Invasive alternatives to IVC

While IVC is the most prevalent invasive means to measure ICP, alternatives exist. These devices typically insert a measurement device at a shallower distance, which carries a lower risk of infection or complication, but reduces accuracy and precision. The methods essentially vary by depth of penetration. Examples include intraparenchymal sensors which are inserted into the brain tissue [Lev77]; subarachnoic screws inserted into the fluid space between the brain and skull [VBY73]; and subdural and epidural sensors inserted in the protective meninx layers of the skull [RSC79]. With the exception of subarachnoic screws, these methods do not allow CSF drainage, and so they are less actionable.

2.1.4 Clinical analysis of ICP

ICP varies with body position, and so reference values are taken from the supine (flat) position in critical care. In this position, a healthy range of mean ICP is around 7–15 mmHg. A reading above 15 mmHg is regarded as elevated; 20 mmHg is abnormal for a TBI patient; and 25 mmHg requires aggressive treatment [CP04]. ICP varies over time, and *mean ICP* is typically taken over longer timescales, such as 30 minutes [CP04].

The periodic dynamics of ICP signals are also used for wavefrom analysis. These analyses range from heuristics that can be clinically calculated to dynamical system modeling. An example heuristic is RAP, the correlation coefficient (R) between the amplitude of the ICP signal at the heart rate frequency (A), and the mean ICP (P) [CP04]. A low RAP value, indicating low correlation, demonstrates a lack of synchronization between the arterial system pumping blood and the pressure of fluid in the skull. This lack of synchronization shows that autoregulation of fluid pressure is adequate. By contract, a high correlation value shows that ICP is strongly influenced by the cardiac cycle of the heart, indicating low compensation ability. In this case, high blood pressure will cause high ICP, which requires intervention. As such, RAP gives a measure of the robustness to patient condition to changes in blood pressure. Model-based analyses will be discussed in detail later in this chapter.

2.2 Noninvasive ICP estimation

Several noninvasive alternatives to ICP monitoring have been proposed. These methods reduce patient risk, and in many cases can be performed without a neurosurgeon present. These key properties may in future extend ICP monitoring to new patient populations that are not currently indicated, including ischemic stroke and hemorrhage patients, or facilitate on-site TBI testing at car accidents or sports venues [KVN12].

However, no noninvasive ICP (NICP) estimators have yet reached the accuracy and precision necessary for clinical acceptance or the standard set by the Association for the Advancement of Medical Instrumentation [Med15]. This standard stipulates that an ICP measurement device must report mean ICP accurately to within ± 2 mmHg for 0–20 mmHg, and $\pm 10\%$ for 20–100 mmHg. NICP methods are however a subject of ongoing research, and are already useful for preliminary screening, to determine if invasive ICP monitoring is necessary.

This section will discuss a key technology for NICP estimation, *transcranial Doppler ultrasonography*, and then review recent model-based approaches. Many other NICP estimation technologies exist, but are omitted as they do not readily interface with the methods presented in this dissertation. These additional technologies include tympanic membrane displacement [TA07]; optic nerve sheath diameter [KSM08]; magnetic resonance imaging [MMS08]; and computed tomography [MBG81].

2.2.1 TCD ultrasonography

Among the most popular NICP methods is the use of transcranial Doppler (TCD) ultrasonography. These devices measure cerebral blood flow velocity (CBFV), the speed of blood flowing through the middle cerebral artery in the brain. TCDs operate by emitting ultrasonic waves at the temple in the head; the waves are reflected off of the flowing blood and are frequency shifted as per the Doppler effect. The echoes are measured by the device, so that CBFV can be calculated [RBA12]. An example of a CBFV waveform is shown in Fig. 2.2.

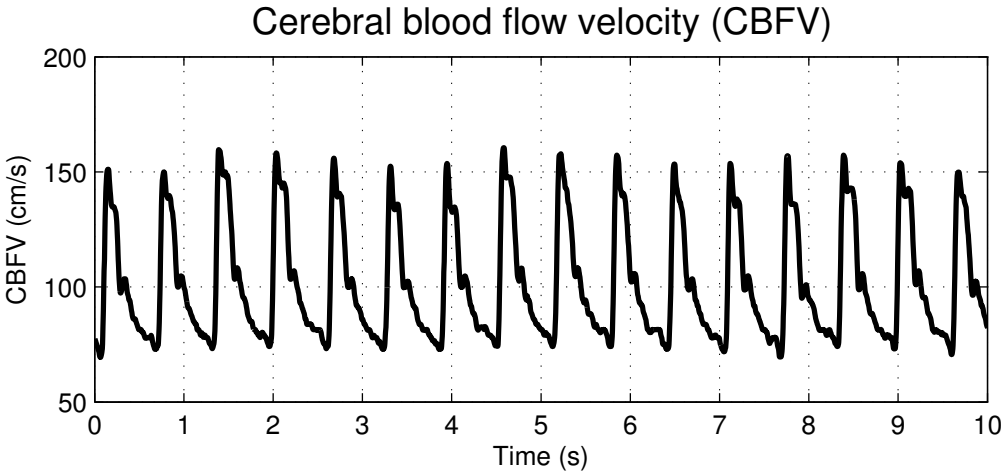

Figure 2.2: A CBFV waveform measured by transcranial Doppler ultrasonography.

A popular clinical heuristic derived from CBFV is the pulsatility index (PI), the ratio of the amplitude of the CBFV wave to its mean value:

$$\text{PI} = \frac{\text{Systolic CBFV} - \text{Diastolic CBFV}}{\text{Mean CBFV}} = \frac{\max(\text{CBFV}) - \min(\text{CBFV})}{\text{avg(CBFV)}}. \tag{2.1}$$

PI is correlated with ICP. Bellner et al. [BRR04] found a strong linear correlation with $R^2 = 0.733$. 95% of points were able to be classified within a range of 4.2 mmHg, for measurements below 30 mmHg. Similarly, Behrens et al. [BLA10] found a 95% confidence interval of 3.8 mmHg in this region. PI has been favored by researchers because it is a ratio, rather than absolute value, removing some dependency on anatomical variation seen in TCD measurements [BRR04]. However, the methods have overall shown variability that make them unsuitable for clinical ICP use [BRR04, BBS10].

TCDs are also subject to noise, large artifacts, and occasional signal dropout from mismeasurement and patient motion [RBN06]. While noise can be filtered out [VSR96], sparse artifacts are more problematic [KL13]. For manually-computed PI-based metrics, these errors can be identified by clinical annotation and ignored. However, the presence of these corruptions is problematic for autonomous methods of TCD analysis.

TCD ultrasonography is noninvasive and easily repeated [BRR04], although 11% of patients cannot provide this measurement at all due to anatomical variability [RBA12]. Overall, TCD estimates can provide reasonably accurate noninvasive estimates of ICP, but they are

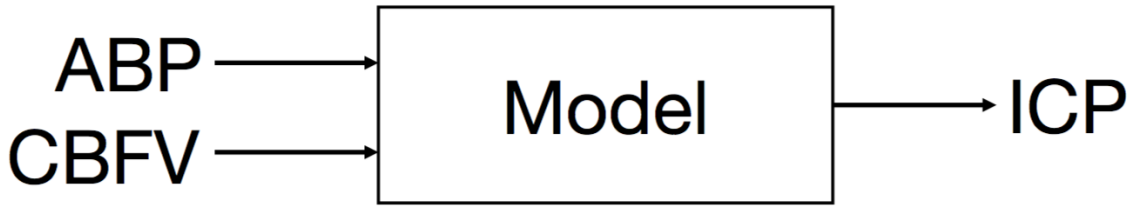

Figure 2.3: A popular input–output configuration for model-based NICP estimators.

not consistent enough to replace invasive methods.

2.2.2 Model-based NICP estimators

Recent literature has sought to replace clinical heuristics for NICP estimation with model-based approaches. Many of these methods rely on physiological input data that can be monitored continuously and can be easily interfaced with bedside computer technology. This integration removes the need for clinicians to manually read data or perform calculations, and allows for real-time estimates of ICP.

Model-based NICP estimators have widely adopted a configuration in which CBFV and ABP are predictive inputs. This black box model is shown in Fig. 2.3. A simple justification for this choice is that CBFV is strongly correlated to ICP, while ABP provides a useful pressure baseline in the body, and is used in CPP calculations. Both signals can be measured continuously, routintely, and noninvasively.

The usage of CBFV and ABP as NICP estimators predates dynamical models. Static relationships based on these signals was studied throughout the 80s and 90s [ALL86, CMS98, CMD92, KCB88]. This work led to the development of early dynamical models with these inputs. Common approaches included compartment flow models [MMW87, UIS95] and analogies to electrical circuits [TKI86].

The three perhaps most widely-cited model-based estimators use CBFV and ABP as their primary inputs. One model by Schmidt et al. [SKS97] proposed a transfer function between

linear transformations of CBFV and ABP to predict ICP. This generalized approach, which does not attempt to describe internal physiology, achieved mean absolute ICP errors around 4.1 mmHg [SKS97, SWS16]. Another successful model by Heldt et al. [KVN12, MFH16] uses an electrical analogy and has a mean (bias) error of 1.6 mmHg under certain conditions.[1]

Finally, a series of models from Hu et al. have also extended dynamical modeling by using machine learning techniques. These approaches include nonlinear [HNG06] and kernel regression [XKB10], data mining mapping functions [HNB06], and semisupervised learning [KHP13]. These methods have shown continuous improvement towards an NICP protocol with median errors below 6 mmHg, but are yet to meet total clinical needs. Neural network–based methods have also been proposed [MMK08].

[1]Note that this quantity is distinct from mean *absolute* error.

CHAPTER 3

Review of subspace system identification

This chapter will review some selected technical topics from the field of *subspace system identification* that are useful for studying linear dynamical systems and their low-order approximations. These topics will underpin the new methods in this dissertation.

The first section provides a structural overview of the systems that will be used. In the second section, the family of subspace system identification methods will be introduced. An important result from this field is highlighted, which relates the *order* of a system to the *rank* of certain data matrices. The third section then discusses an efficient convex heuristic for approximately minimizing over the rank of a matrix.

3.1 Linear time-invariant systems

In this work, ICP dynamics will be modeled by *linear time-invariant* (LTI) systems, a subset of linear dynamical systems. LTI systems have different possible representations; the most widely-used is the *state space* representation [FPE94]:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k), \end{aligned} \tag{3.1}$$

where $u(k) \in \mathbb{R}^m$ is the input signal,[1] $y(k) \in \mathbb{R}^p$ is the output signal, $x(k) \in \mathbb{R}^n$ is the unobserved *state vector*, and (A, B, C, D) are the matrices that define the system. The system is sampled at $k = \{0, 1, 2, \ldots\}$. The dimension of the state vector n is called the *order* of the system, and is a measure of its complexity.

[1] In most practical cases, $u(k)$ will actually be a collection of multiple, scalar (1-dimensional) signals. However, for the sake of analysis it is easier to treat the inputs as a single, multi-variable signal. So, 'input' and 'inputs' will be used interchangeably based on context. This same idea also applies to the output, $y(k)$.

State space structures are well-studied and have numerous advantages [OY02]. However, they require signals to be *labeled* as inputs and outputs. This partitioning has its roots in control systems theory, where inputs are classified as the signals chosen directly by the user [FPE94]. In physiological models, there is no clear notion of a system 'input'; while some labels can be argued from a physiological standpoint (such as heartbeats as inputs), others can be ambiguous. For these cases, it can be useful to consider a more general system representation of the form

$$R_0 w(k) + R_1 w(k+1) + \cdots + R_\ell w(k+\ell) = 0, \tag{3.2}$$

where $w(k) \in \mathbb{R}^d$ is a signal for which there is a linear relationship between up to $\ell + 1$ consecutive observations [MWV06]. The matrices R_i define the system.

This general representation is much less useful in practical settings, and will not be used explicitly. However, all state space systems can be represented in this form and, in turn, for every LTI system of form (3.2) there is a *causal* partitioning of $w(k)$ into input and output signals [MWV06]:

$$w(k) = P^\top \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}. \tag{3.3}$$

The relationship between these two representations is useful when results for a state space system can be extended to cases where data can be unlabeled.

3.2 Subspace system identification

System identification is the process of fitting models of dynamical systems to data. There are two broad classes of methods to solve this problem. The first are *prediction error minimization* methods, which operate by minimizing some function of model error, such as the difference between the measured signals and those produced by a simulation of the model [Lju98, Pel15]. These methods will not be used here. *Subspace system identification*, which is leveraged in this dissertation, is a fundamentally different approach in which input–output data is used to construct matrices with linear algebraic properties that reveal the system structure [Lju98].

This section will review some key principles of subspace methods. This includes a basic overview of how they are used to identify systems from data, and a review of an important result on system order. As this dissertation will only use a subset of the theory from this field, it is not a comprehensive review; interested readers can refer to [VV07, VD12, Lju98] for further details.

A note on block Hankel matrix notation. The upcoming sections will frequently use block Hankel matrices constructed from a sequence of vectors. Such a matrix H with r columns that contains N vector blocks $h(\cdot) \in \mathbb{R}^d$, starting at $h(0)$ through $h(N-1)$, will be labeled $H_{r,N}$. That is,

$$H_{r,N} = \begin{bmatrix} h(0) & h(1) & \cdots & h(r-1) \\ h(1) & h(2) & \cdots & h(r) \\ h(2) & h(3) & \cdots & h(r+1) \\ \vdots & \vdots & \ddots & \vdots \\ h(N-r) & h(N-r+1) & \cdots & h(N-1) \end{bmatrix} \in \mathbb{R}^{(N-r+1)d \times r}. \tag{3.4}$$

3.2.1 Overview of subspace identification process

Subspace identification methods retrieve approximations of the system matrices (A, B, C, D) and the initial system state $x(0)$. This is done in two steps. First, the observability matrix $\mathcal{O}_r$ of the system is estimated. This reveals the matrices A and C. Then, system equations are linear in the remaining unknowns, B, D, and $x(0)$, and so these can be solved using least squares.

Classically, the derivation of subspace methods begins by noting that the state space equations (3.1) can be rearranged to give the *input–output equation*:

$$Y_{r,N} = \mathcal{O}_r X_{N,N} + \mathcal{T}_r U_{r,N}, \tag{3.5}$$

where $\mathcal{O}_r$ and $\mathcal{T}_r$ are the matrices

$$\mathcal{O}_r = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{r-1} \end{bmatrix}, \qquad \mathcal{T}_r = \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots \\ CA^{r-1}B & CA^{r-3}B & CA^{r-4}B & \cdots & D \end{bmatrix}, \tag{3.6}$$

and $U_{r,N}$, $Y_{r,N}$ and $X_{N,N}$ are block Hankel matrices of $u(k)$, $y(k)$, and $x(k)$ using the notation in (3.4). Note that $\mathcal{O}_r$ is the *observability matrix* of the system.

The aim of the system identification task is to estimate $\mathcal{O}_r$ and $\mathcal{T}_r$ in (3.5)—which contain the system matrices—using $Y_{r,N}$ and $U_{r,N}$ from the measurement data. This problem is nontrivial because $X_{N,N}$ is also unknown. So, several methods exist to first estimate $\mathcal{O}_r$.

In one short example, the term $\mathcal{T}_r U_{r,N}$ is removed from (3.5) by applying $\Pi^{\perp}_{U_{r,N}}$, the projection onto the nullspace of $U_{r,N}$, to both sides of the equation:

$$Y_{r,N}\Pi^{\perp}_{U_{r,N}} = \mathcal{O}_r X_{N,N}\Pi^{\perp}_{U_{r,N}}. \tag{3.7}$$

(This projection can be computed from $U_{r,N}$ using a singular value decomposition; see [Lau05] for details).

The next steps of analysis assume that the following condition holds [LHV13]:

$$\operatorname{rank}\left(\begin{bmatrix} X_{N,N} \\ U_{r,N} \end{bmatrix}\right) = n + \operatorname{rank}(U_{r,N}). \tag{3.8}$$

This condition is called the *persistency of excitation of a sufficient order*, and can be thought of as the requirement for the input $u(k)$ to be sufficiently complex as to excite all of the modes of the system. Because $X_{N,N}$ and n are not known, the condition cannot be verified, although it is realistic for real input data and low-order systems. See [VV07, §10.2.4.2] and [JW98] for further details.

If (3.8) holds, then from (3.7) it follows that

$$\operatorname{rank}(Y_{r,N}\Pi^{\perp}_{U_{r,N}}) = n, \tag{3.9}$$

and

$$\begin{aligned} \text{range}(Y_{r,N}\Pi_{U_{r,N}}^{\perp}) &= \text{range}(\mathcal{O}_r X_{N,N}\Pi_{U_{r,N}}^{\perp}) \\ &= \text{range}(\mathcal{O}_r). \end{aligned} \tag{3.10}$$

The first result (3.9) states that the rank of $Y_{r,N}\Pi_{U_{r,N}}^{\perp}$ is equal to the order of the system [VV07]. When the data is corrupted by noise, this rank deficiency does not perfectly occur. In this scenario, a low-rank approximation of $Y_{r,N}\Pi_{U_{r,N}}^{\perp}$ can be found by taking its SVD and truncating its singular values, which produces a low-order approximation of the system [LHV13].

The second result (3.10) demonstrates that the the column space (or *range*) of $\mathcal{O}_r$ can be determined from the data matrix $Y_{r,N}\Pi_{U_{r,N}}^{\perp}$. Because LTI system matrices are unique up to a similarity transform [OY02], any basis of this column space is suitable as a realization of $\mathcal{O}_r$. (One way to realize this matrix is to take the SVD of $Y_{r,N}\Pi_{U_{r,N}}^{\perp}$; again, see [Lau05] for details).

Once $\mathcal{O}_r$ and hence A and C are known, the input–output equation (3.5) is linear in its unknowns. So, B, D, and $x(0)$ can be estimated using least squares [BV18].

There are many variations of subspace methods, which largely differ in how $\mathcal{O}_r$ is calculated. Examples include N4SID, MOESP, IVC, CVA, and, most recently, N2SID; details can be found in [VD94, Lju98, VH15, VH16].

3.2.2 Data matrices that reveal system order

Equation (3.9) is an important result: it demonstrates that data matrices can be designed whose *rank* provides insight on the the *order* of the system. A consequence of this idea is that a low-rank estimate of such a matrix will provide a low-order approximation of this system.

Another data matrix with a similar property is

$$F(u,y) = \begin{bmatrix} U_{r,N} \\ Y_{r,N} \end{bmatrix}, \tag{3.11}$$

which is simply the block Hankel matrix of inputs, $U_{r,N}$ stacked above the block Hankel matrix of outputs $Y_{r,N}$. If (3.8) holds then [MDV89, MV90]:

$$\operatorname{rank}(F(u,y)) = n + \operatorname{rank}(U_{r,N}). \tag{3.12}$$

If $U_{r,N}$ is full row rank then this becomes

$$\operatorname{rank}(F(u,y)) = n + m(N-r+1), \tag{3.13}$$

and so the rank of $F(u,y)$ is the system order, plus a constant.

Both $Y_{r,N}\Pi_{U_{r,N}}^{\perp}$ and $F(u,y)$ are interesting to consider as candidates for a rank/order minimization problem. Of these two choices, however, $F(u,y)$ has two distinct benefits. Its most immediate advantage is that it is linear in both $u(k)$ and $y(k)$, so all of the measured signals can be treated as variables in an estimation problem with ease.

Another benefit to using $F(u,y)$ is perhaps less obvious. Consider the unlabeled signal format

$$w(k) = P^{\top} \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}, \tag{3.14}$$

where P is again the permutation that partitions the unordered signals into inputs and outputs. Then, the block Hankel matrix

$$W_{r,N} = \begin{bmatrix} w(0) & w(1) & \cdots & w(r-1) \\ w(1) & w(2) & \cdots & w(r) \\ w(2) & w(3) & \cdots & w(r+1) \\ \vdots & \vdots & \ddots & \vdots \\ w(N-r) & w(N-r+1) & \cdots & w(N-1) \end{bmatrix} \tag{3.15}$$

is simply a row permutation of $F(u,y)$. Because rank is a unitarily invariant function,

$$\begin{aligned} \operatorname{rank}(W_{r,N}) &= \operatorname{rank}(F(u,y)) \\ &= n + m(N-r+1) \end{aligned} \tag{3.16}$$

when $U_{r,N}$ is full rank. So, when the input–output partitioning of the signals is not known or defined, there is still a relationship between the rank of the data matrix $W_{r,N}$ and the

order of the system. While the actual rank cannot be determined without choosing m, the dimension of the input signal, there is still a *monotonically increasing* relationship between the rank of $W_{r,N}$ and n. So, a low-rank approximation of $W_{r,N}$ should still produce low-order-inducing signals, even if the inputs and outputs of the system are not explicitly declared. This more relaxed view is useful for physiological models in which input–output partitioning is ambiguous.

3.3 Trace norm minimization

An important takeaway from Section 3.2 is that in certain instances, the rank of a data matrix can infer the order of an LTI system. So, the task of finding *low-order* approximations of a system is closely related to that of finding *low-rank* approximations of a matrix.

However, it is difficult to compute low-rank approximations of matrices in many cases—for example in the presence of constraints on its structure—because rank minimization is a nonconvex problem [Faz02]. A popular alternative is *trace norm minimization* [FHB01]. The trace norm (or *nuclear norm*) $\|X\|_*$ of a matrix $X \in \mathbb{R}^{m\times n}$ is the sum of its singular values:

$$\|X\|_* = \|\sigma\|_1 = \sum_{i=1}^{\min(m,n)} \sigma_i, \qquad \text{for the SVD } X = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^\top. \tag{3.17}$$

It is the tightest convex relaxation of rank over the unit spectral norm ball [Faz02]. Applying trace norm minimization to a matrix is equivalent to applying ℓ_1-norm to the vector of its singular values. This process tends to produce a sparse set of singular values, and, in turn, a low-rank matrix. Minimizing the trace norm of a matrix, subject to constraints, is a very effective, indirect heuristic for finding a low-rank solution to a problem.

Techniques exist to improve upon the trace norm relaxation. Grussler et al. [GRG18, GR15] present an approach to solving rank-constrained optimization problems using a convex relaxation with certain verifiable guarantees on rank, ensuring a low-rank solution. This approach is useful when the desired rank of the variable matrix is known. Another simple technique is to iteratively re-weight the terms in the norm [MF10].

3.3.1 Applications of trace norm minimization

System identification. There are several instances of trace norm minimization being applied to subspace system identification. A common theme in these methods is to replace the rank minimization step, which is traditionally done by truncating the singular values of a data matrix [VV07], with a convex optimization problem.

For example, Liu & Vandenberghe [LV09] proposed solving the convex problem

$$\text{minimize} \qquad \|Y_{r,N}\Pi_{U_{r,N}}^{\perp}\|_* + \lambda \sum_{k\in T} \|y(k) - y_{\text{meas}}(k)\|_2^2 \tag{3.18}$$

to identify low-rank systems. The problem computes $y(k)$, an approximation of measurements $y_{\text{meas}}(k)$ that typically produces a low-order system. A system realization can then be quickly found by applying the results (3.9)–(3.10). While (3.18) is similar in effect to truncating the singular values of $Y_{r,N}\Pi_{U_{r,N}}^{\perp}$, its convex formulation is much more flexible. For example, it is allowable for $y_{\text{meas}}(k)$ to have missing elements. Another variation [LHV13] is

$$\text{minimize} \quad \|F(u,y)\|_* + \lambda \sum_{k\in T_0} \|y(k) - y_{\text{meas}}(k)\|_2^2 + \gamma \sum_{k\in T_1} \|u(k) - u_{\text{meas}}(k)\|_2^2, \tag{3.19}$$

where missing input data $u_{\text{meas}}(k)$ is also allowable. These methods have been successfully applied to data from the DaISy benchmark dataset [DDD97] to identify LTI models with some missing data.

Verhaegen & Hansson [VH14, VH16] present a similar method that minimizes the trace norm of a variation on the input–output equation (3.5). Smith [Smi12] uses trace norm minimization to perform subspace system identification in the frequency domain instead of the time domain. Sznaier & Ayazoglu [AS12, SAI14] apply the idea to robust system identification, to find low-order system estimates with guarantees on stability and error bounds.

Other applications. There are numerous applications for trace norm minimization beyond system identification. Candès & Plan overview in [CP10] its role in matrix completion and collaborative filtering, where a data matrix can be recovered using only a small sample

of its entries. An example of its use in robust principal component analysis is provided in [XCS10], where it is used to recover a low-dimensional subspace in the data that is expressed as a low-rank matrix. Applications for trace norm minimization also appear in computer vision, where low-rank matrices can model the background in a video sequence [CLM11], or used to estimate the motion of an object when it is temporarily obscured [AYS13].

CHAPTER 4

Enhancing corrupted physiological signals by trace norm regularization

Note: A variation of this chapter has been submitted to IEEE Transactions on Biomedical Engineering as *Artifact Rejection and Missing Data Imputation in Cerebral Blood Flow Velocity Signals via Trace Norm Minimization* (Gunn, C.A., Hu, X., Vandenberghe, L.). An earlier version was presented at the 16th International Symposium of Intracranial Pressure and Neuromonitoring as *A Novel Signal and Model Enhancement Algorithm Improves Simulation of ICP Signals using CBFV and ABP Signals* (Gunn, C.A., Vandenberghe, L., Hu, X.).

4.1 Introduction

4.1.1 Motivation

Many physiological signals are degraded by significant corruptions that limit their clinical usefulness. An example that is pertinent to NICP estimation is the use of CBFV signals that are measured by transcranial Doppler ultrasonography. The TCD measurement process uses ultrasonic probes that are either placed by hand or fixed on an elastic headframe that is fit around the skull [Aas12]. Patient motion and clinical disruptions can cause relative movement or even dislodgement of these probes, introducing large artifacts that dominate the desired signal in the measurement, or sections of entirely missing data [RBN06]. Some examples of these errors are shown in Fig. 4.1. In traditional clinical settings these corruptions can be problematic, but generally can be manually identified and ignored. In automated

settings, such as the model-based NICP estimators described in Section 2.2.2, there may be no human intervention step, and the effect of these corruptions on ICP estimates may not be obvious.

This chapter will present a suite of methods to remove artifacts and impute sections of missing data in corrupted CBFV signals. The methods exploit the knowledge that the relationship between CBFV, ABP, and ICP can be well-approximated by low-order LTI systems. Measured signals are enhanced by fitting them to low-order dynamical models, using convex regularization terms that improve robustness to large deviations and missing data. The methods are based on convex optimization formulations, and utilize the results outlined in Chapter 3 in subspace system identification and trace norm minimization. Simulations demonstrate that the method successfully removes real CBFV artifacts, and can impute missing data of with reasonable accuracy. The methods are presented in a general context, and are applicable to other signal processing applications that meet the necessary assumptions.

4.1.2 Background on current methods

The tasks of artifact rejection and missing data imputation are in general distinct. Artifact rejection is the task of removing *artifacts* from a signal, which are here defined as measurement errors that are large, sparse, and often occur in groups of successive measurements [SPH08]. Missing data imputation is simply the process of estimating missing elements from an ordered data set, such as time series data [Efr94].

Artifact rejection. Popular mechanisms for artifact rejection include smoothing and statistical approaches. Smoothing-type approaches assume that artifacts are equivalent to sharp signal transitions, and attempt to remove them by smoothing the signal. Examples include median filtering [RL05] and sparse quadratic smoothing [BV04]. These methods are computationally efficient and require little information. However, the assumption that all sharp transitions are erroneous measurements is often unrealistic. For example, many physiological signals, including CBFV, are driven by heartbeats and experience sharp changes at the be-

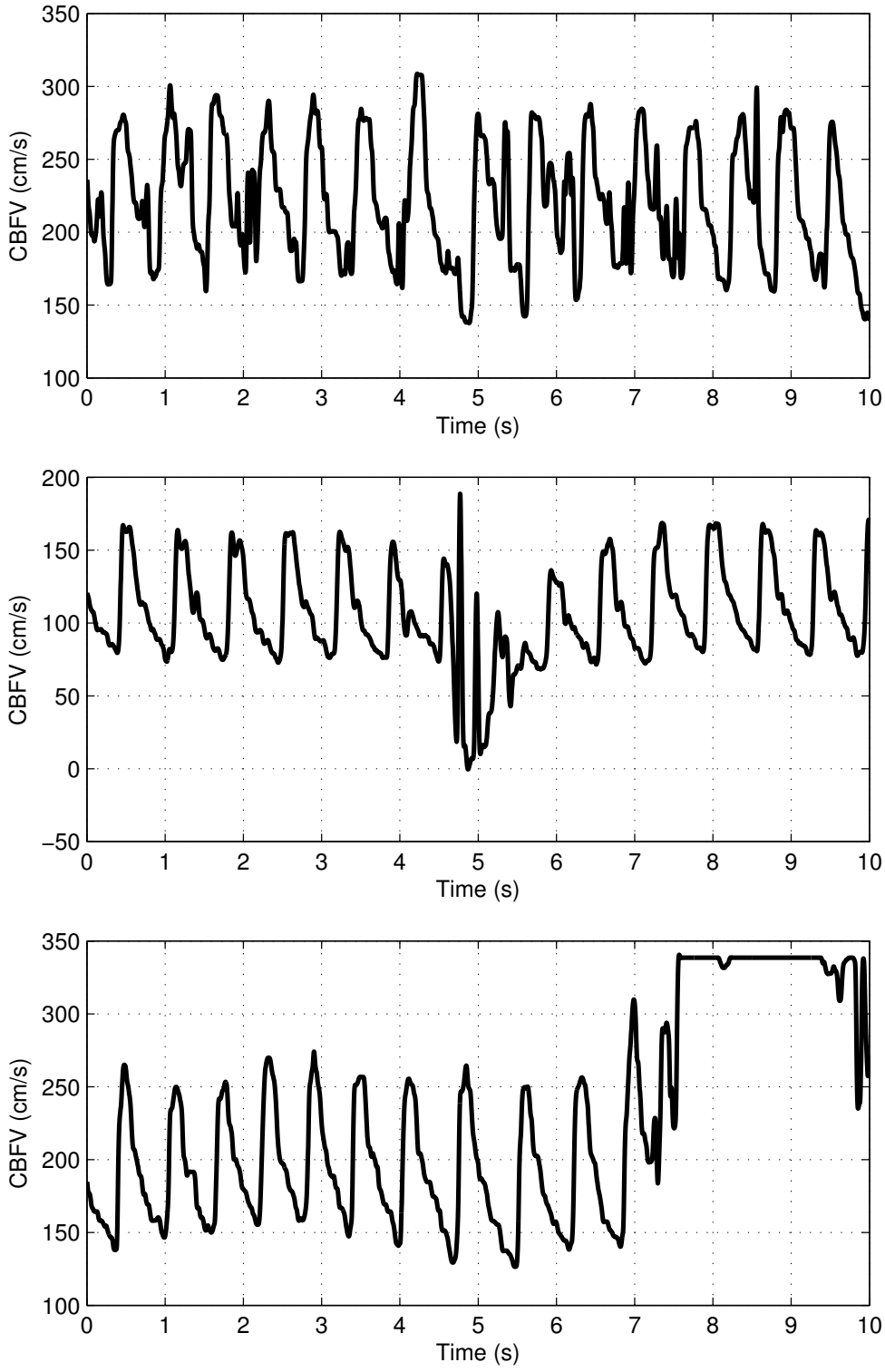

Figure 4.1: Examples of CBFV waveforms with various types of large artifacts.

ginning of systole [Hal15] (see, for example, the clean sections of Fig. 4.1). Other approaches that are more statistically-driven include statistical thresholding [NWR10] and independent component analysis [JHL98]. These particular methods have been successfully been applied to physiological applications. However, they rely on stationary data and are not robust to systematic shifts that may occur across a set of related signals.

The methods produced in this dissertation are distinct from the above approaches by their strong reliance on system dynamics. They assume that artifacts are *sparse events that increase the complexity of dynamics between signals in a low-order system*. This approach notes that if multiple signals that form a low-order dynamical system are measured together, then the presence of a corruption in any of the signals will introduce unexplained, complex dynamics to this relationship. This approach is robust to many limitations of the previous methods because it compares behavior between signals to differentiate artifacts from system-wide external events.

Missing data imputation. Current tools for missing data imputation are largely suitable for physiological signal processing applications. Given a set of signals that are related by a linear dynamical system, Kalman filters can be used to estimate the unobserved state sequence and then reconstruct the signals [SSF04, KA86]. This approach relies on the dynamics between signals, although it requires user selection of a model. Approaches without a foundation in dynamics, such as interpolation and multiple imputation [KHJ01], should be avoided in these contexts.

4.2 Model validation

A key assumption in this dissertation is that the relationship between ABP, CBFV, and ICP can be accurately described by a low-order LTI system. This section verifies this assumption with data. First, the physiological data used will be described. Then, two LTI models are proposed that will be frequently used: the *ICP model* and the *CBFV model*. Finally, the accuracy of these models is validated.

4.2.1 ICP data

Analysis was performed using retrospective clinical data collected for the previous studies reported in [KSB12, KHP13]. The dataset contains 75 patients, recorded at the University of California, Los Angeles (UCLA) Medical Center. 38 patients had TBI, 27 had aneurysmal subarachnoid haemorrhage (aSAH), and 10 had normal pressure hydrocephalus (NPH). Data was separated into a total of 169 episodes, each 360 heartbeats in length. CBFV, ABP, and ICP signals were recorded at 400 Hz, and downsampled to 40 Hz. CBFV was measured noninvasively via transcranial Doppler. ABP was measured routinely by catheter.[1] ICP was measured invasively by IVC. Patients had a median age of 34 with interquartile range (IQR) 24–56, and a total range of 18–89.

4.2.2 Model configurations

Two related clinical model configurations will be used: the *CBFV model*, and the *ICP model*. Both cases arise in TBI care. Their configurations are shown in Fig. 4.2. The CBFV model uses ABP as its input, and CBFV as its output:

$$u(k) = \mathrm{ABP}(k) \in \mathbb{R}, \qquad y(k) = \mathrm{CBFV}(k) \in \mathbb{R}. \tag{4.1}$$

As both of these quantities are routinely measured in critical care, this model is readily computable. An example of such an LTI model is analyzed in [KCS98].

The ICP model uses ABP and CBFV as inputs, and ICP as its output:

$$u(k) = \begin{bmatrix} \mathrm{ABP}(k) \\ \mathrm{CBFV}(k) \end{bmatrix} \in \mathbb{R}^2, \qquad y(k) = \mathrm{ICP}(k) \in \mathbb{R}. \tag{4.2}$$

This configuration is useful when a patient's ICP is measured invasively via IVC. Examples of LTI systems with this configuration can be found in [SKS97, KVN12].

[1] While this process is technically invasive, its measurement is standard practice in intensive care [WL07], and for simplicity will be classified as 'noninvasive' in the sense that it does not require the patient undergo an *additional* invasive procedure for the information to be available.

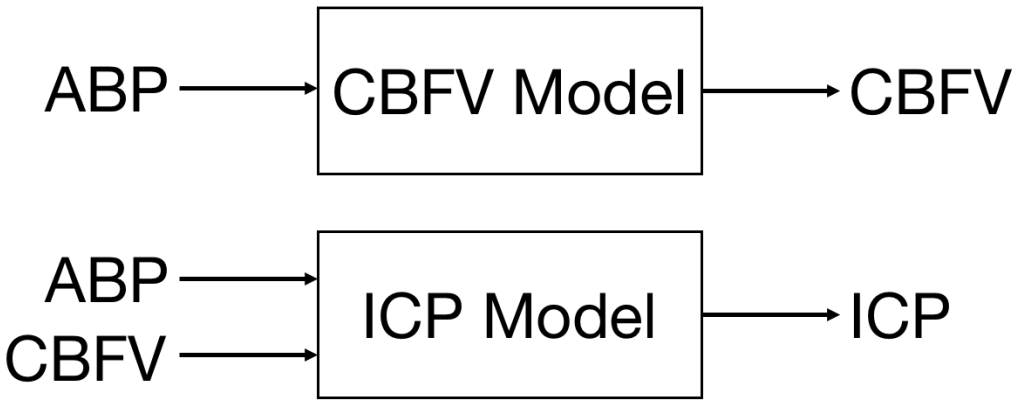

Figure 4.2: Input–output configurations of the CBFV and ICP models.

4.2.3 Simulation and results

LTI models were fit to training data, and validated on testing data. The following process was repeated for each of the 169 episodes, for both the CBFV model and the ICP model. First, the signal data was split into a training sequence and a testing sequence, each equal in length. A *patient-specific* LTI model was fit to the training data using the N4SID method (as implemeneted in [Lju07]). Then, the input data from the testing sequence was applied to the model, which produced a simulated output sequence. This simulation was compared to the true, measured output of the testing sequence. An example of the simulation output of both model configurations is provided in Fig. 4.3.

The ICP model had a median [IQR] absolute error of 1.02 [0.46–2.10] mmHg (1.13%). These simulations compare favorably to the ANSI/AAMI standard for measurement devices ($\pm$2.0 mmHg for 0–20 mmHg, and $\pm$10% in the range 20–100 mmHg [Med15]). For the CBFV model, the median [IQR] absolute error was 6.57 [2.85–14.49] cm/s (7.38%).

Additionally, *population-wide* models were also constructed by fitting a model over the entire training set, and validating both over the entire testing set. These models performed poorly, with median absolute errors of 5.44% and 32.74% for the ICP and CBFV models respectively. Results of these simulations are summarized in Table 4.1.

These results demonstrate that the *patient-specific ICP model* is an excellent choice for

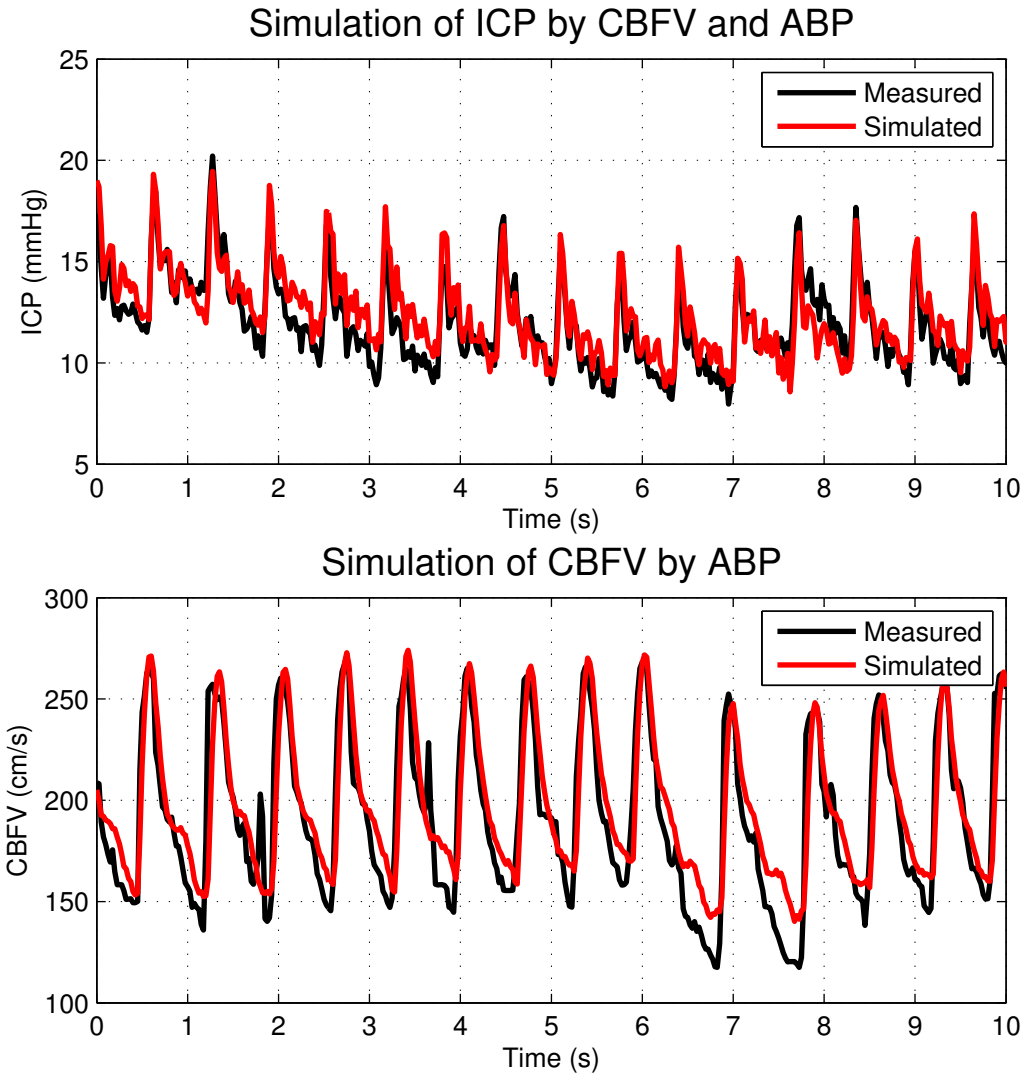

Figure 4.3: Simulated output of the ICP model and the CBFV model in two separate episodes.

Table 4.1: Simulation error the CBFV and ICP models, for both patient-specific and population-wide modeling cases. Errors are given in median absolute error, expressed as a percentage.

Model data	ICP model (%)	CBFV model (%)
Patient-specific	1.13	7.38
Population-wide	5.44	32.74

modeling ICP dynamics. While the patient-specific CBFV model gave inferior performance, its error was reasonable for a single-input-single-output LTI model of a complex physiological process. These results validate the assumption that the data in this dissertation can be accurately modeled by LTI systems. For both configurations, population-wide models substantially underperformed, so these will be mostly ignored.

The results also suggest an upper-limit on future performance goals. For example, if an ICP model is constructed without knowing a patient's ICP—because it was not measured—then the resulting model should underperform the results in the above experiment, specifically, a median [IQR] absolute error of 1.02 [0.46–2.10] mmHg. Hence, this result serves as a performance benchmark for NICP estimation methods that use LTI models.

4.3 Methods for artifact rejection and missing data imputation

4.3.1 Basic method

Consider a set of signal measurements $w_{\text{meas}}(k) \in \mathbb{R}^d$, with $k = 0, \ldots, N-1$. If the signals are labeled as inputs $u_{\text{meas}}(k)$ and $y_{\text{meas}}(k)$ then

$$w_{\text{meas}}(k) = \begin{bmatrix} u_{\text{meas}}(k) \\ y_{\text{meas}}(k) \end{bmatrix}. \tag{4.3}$$

The signals may contain artifacts or sections of missing data. Elements that were measured are indexed by the index set $\mathcal{K}_i \subseteq \{0, \ldots, N-1\}$, where $i = 1, \ldots, d$. It is assumed that the input data $u_{\text{meas}}(k)$ is persistently exciting of sufficient order as to satisfy the rank condition (3.8) (even if the input–output labels of $w_{\text{meas}}(k)$ are not explicitly known). If all of the measurements are available for the ith signal, it can notated as $w_{i,\text{meas}} \in \mathbb{R}^N$.

The aim of this method is to determine an optimized signal $w(k) \in \mathbb{R}^d$ with artifacts removed and missing data imputed. The signal $w(k)$ will be the optimization variable in the method. The cost function combines two objectives. The first is to minimize $\|W_{r,N}\|_*$, as a penalty on values of $w(k)$ that increase the order of the model dynamics. The second term penalizes deviation of $w(k)$ from the measured data. The tradeoff between the terms

is controlled by regularization parameters $\lambda_i > 0$ that correspond to each signal i. Hence, the following convex optimization problem is posed:

$$\text{minimize} \qquad \|W_{r,N}\|_* + \sum_{i=1}^{d} \sum_{k \in \mathcal{K}_i} \lambda_i |w_i(k) - w_{i,\text{meas}}(k)|. \tag{4.4}$$

When there is no data missing this problem becomes

$$\text{minimize} \qquad \|W_{r,N}\|_* + \sum_{i=1}^{d} \lambda_i \|w_i - w_{i,\text{meas}}\|_1. \tag{4.5}$$

The second term in both (4.4) and (4.5) are weighted ℓ_1-norm data-dissimilarity penalties, and are chosen for their robustness against outliers. They allow for large corrections to be made to the signal that reasonably reduce $\|W_{r,N}\|_*$, and so provide the desired artifact rejection property.

When imputing missing data is the only goal, without modifying the the measured signals, the following constrained formulation can be used:

$$\begin{aligned} \text{minimize} \qquad & \|W_{r,N}\|_* \\ \text{subject to} \qquad & w_i(k) = w_{i,\text{meas}}(k), \quad i = 1, \ldots, d, \quad k \in \mathcal{K}_i. \end{aligned} \tag{4.6}$$

This problem can be interpreted as a structured *low-rank matrix completion problem* [RFP10], that is, the missing entries that comprise the block Hankel matrix $W_{r,N}$ are completed.

4.3.2 Structured artifact rejection

A limitation of the method expressed in (4.4) is that it penalizes each residual element $w_i(k) - w_{i,\text{meas}}(k)$ independently, without respect to their ordering. In practice, however, artifacts occur in blocks of successive measurements. These block artifacts can be removed using overlapping group sparsity penalties. Such penalties have been proposed by Yuan and Lin [YL06], Obozinski et al. [OJV11], and others [PF11, BJM12, KX10].

The structured blocks in question may be just a few measurements in length, each representing a singular artifact. Alternatively, they may be as long as several minutes, representing periods of care in which there was clinical disruption. This choice is left to the user, based on their needs.

To simplify discussion, this section will assume that only one signal $l \in \{1, \ldots, d\}$ (specifically, CBFV) is corrupted by artifacts, and that it is complete. Measurements of the other signals are assumed to be exact, although may be missing. By the approaches of (4.4)–(4.6), a good choice of signal enhancement method for this scenario would be

$$\begin{aligned} \text{minimize} \qquad & \|W_{r,N}\|_* + \lambda \sum_{k=0}^{N-1} |w_l(k) - w_{l,\text{meas}}(k)| \\ \text{subject to} \qquad & w_i(k) = w_{i,\text{meas}}(k),\ i \neq l, k \in \mathcal{K}_i. \end{aligned} \tag{4.7}$$

The aim of this approach is to replace the element-wise penalty on $w_l(k) - w_{l,\text{meas}}(k)$ with an overlapping group sparsity penalty.

Consider a scalar signal $x \in \mathbb{R}^N$, and its measurement x_{meas} which is corrupted by block artifacts of maximum length q. The aim is to find a residual vector $e = x - x_{\text{meas}}$ that can be written as a sum of a small number of contiguous nonzero sections (or *atoms*) of length q. This is written as

$$e = Pz = \begin{bmatrix} P^0 & \cdots & P^{N-q} \end{bmatrix} \begin{bmatrix} z(0) \\ \vdots \\ z(N-q) \end{bmatrix}, \tag{4.8}$$

where each atom $z(k)$ is a signal of length q. Each block $P^k \in \mathbb{R}^{N \times q}$ maps the atom $z(k)$ to positions $k, \ldots, k+q-1$ in the residual e, forming an overlapping structure. These overlapping blocks are then combined to yield e. For example, when $q = 2$,

$$P = \left[\begin{array}{cc|cc|c|cc|cc} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{array}\right]. \tag{4.9}$$

The number of atoms in this representation can be made small by minimizing an ℓ_1/ℓ_2-norm penalty on the set of each $z(k)$. This motivates the overlapping group sparsity penalty

function

$$\Omega(e) = \min_{z} \left\{ \sum_{k=0}^{N-q} \|z(k)\|_2 \middle| e = Pz \right\}. \tag{4.10}$$

Replacing the element-wise penalty of (4.7) with Ω, it becomes

$$\begin{aligned} \text{minimize} \quad & \|W_{r,N}\|_* + \lambda \Omega(w^l - w^l_{\text{meas}}) \\ \text{subject to} \quad & w_i(k) = w_{i,\text{meas}}(k), \quad i \neq l, k \in \mathcal{K}_i. \end{aligned} \tag{4.11}$$

To handle the function Ω algorithmically, the problem is expanded and auxiliary variables $z(k)$ are introduced to obtain

$$\begin{aligned} \text{minimize} \quad & \|W_{r,N}\|_* + \lambda \sum_{k=0}^{N-q} \|z(k)\|_2 \\ \text{subject to} \quad & w^l - w^l_{\text{meas}} = Pz \\ & w_i(k) = w_{i,\text{meas}}(k), \quad i \neq l, k \in \mathcal{K}_i. \end{aligned} \tag{4.12}$$

It is worth noting that this method is not restricted to the simple shifting structure of artifact groups that has been suggested. In fact, arbitrary groups can be defined by the user by choice of the matrix P. One systematic choice that users may also find useful is to penalize every contiguous sequence *up to* length q, rather than of exact length q. Alternatively, groups with length $q_{\min} \leq q \leq q_{\max}$ could be used.

4.3.3 Data-dissimilarity penalties

Because (4.4) is framed as convex optimization problem, it is easy to extend it to new cases by introducing penalties and constraints to capture prior information or different scenarios. For example, the artifact rejection property can be replaced with noise rejection by instead solving

$$\text{minimize} \quad \|W_{r,N}\|_* + \sum_{i=1}^{d} \sum_{k \in \mathcal{K}_i} \frac{\lambda_i}{2} \left(w_i(k) - w_{i,\text{meas}}(k) \right)^2, \tag{4.13}$$

or, for data without missing elements,

$$\text{minimize} \quad \|W_{r,N}\|_* + \sum_{i=1}^{d} \frac{\lambda_i}{2} \|w_i - w_{i,\text{meas}}\|_2^2. \tag{4.14}$$

In fact, the data-dissimilarity penalty on $w_i - w_{i,\text{meas}}$ for each signal i can be uniquely chosen. This is useful if one signal (e.g. CBFV) is corrupted by artifacts, while other signals (e.g. ABP and ICP) are affected by noise, and so different penalties are appropriate. In addition to the ℓ_1- and ℓ_2-norms, another useful choice is the Huber loss function [Hub64].

4.3.4 Smoothing penalties

Users are also able to add smoothing penalties to $w(k)$ if desired. Two common choices of smoothing penalties are *quadratic smoothing* [BV04] and *total variation regularization* [ROF92]. For complete signal data corrupted by artifacts, quadratic smoothing can be achieved by solving

$$\text{min.} \quad \|W_{r,N}\|_* + \sum_{i=1}^{d} \lambda_i \|w_i(k) - w_{i,\text{meas}}(k)\|_1 + \sum_{i=1}^{d} \sum_{k=1}^{N-1} \frac{\gamma_i}{2} \left(w_i(k) - w_i(k-1)\right)^2, \tag{4.15}$$

and, total variation regularization by

$$\text{min.} \quad \|W_{r,N}\|_* + \sum_{i=1}^{d} \lambda_i \|w_i(k) - w_{i,\text{meas}}(k)\|_1 + \sum_{i=1}^{d} \sum_{k=1}^{N-1} \gamma_i |w_i(k) - w_i(k-1)|. \tag{4.16}$$

In both cases $\gamma_i > 0$ are constants controlling the amount of smoothing applied to signal i. Both methods penalize the rate of change of the signal via $w_i(k) - w_i(k-1)$.

4.3.5 Exploiting prior information

Prior information on corruptions can be exploited by choice of penalties and weights. One such case is when the sections of corrupted signal are known a priori. If the clean elements of $w_i(k)$ are described by the index set $\mathcal{C}_i \subseteq \{0, \ldots, N-1\}$, and the data is complete, then it is appropriate to fix different weights for these measurements:

$$\begin{aligned} \text{minimize} \quad \|W_{r,N}\|_* &+ \sum_{i=1}^{d} \sum_{k \in \mathcal{C}_i} \lambda_i^{\text{clean}} |w_i(k) - w_{i,\text{meas}}(k)| \\ &+ \sum_{i=1}^{d} \sum_{k \notin \mathcal{C}_i} \lambda_i^{\text{corrupt}} |w_i(k) - w_{i,\text{meas}}(k)|, \end{aligned} \tag{4.17}$$

with $\lambda_i^{\text{clean}} > 0$ and $\lambda_i^{\text{corrupt}} > 0$ the weights applied to the clean and corrupt data in the ith signal respectively. In an extreme case, each term $w_i(k) - w_{i,\text{meas}}(k)$ can be weighted

individually by some $\lambda_i(k) \in \mathbb{R}$, with $k = 0, \ldots, N-1$.

A constrained variation of (4.17) is

$$\begin{aligned} \text{minimize} \quad & \|W_{r,N}\|_* + \sum_{i=1}^{d} \sum_{k \notin \mathcal{C}_i} \lambda_i |w_i(k) - w_{i,\text{meas}}(k)| \\ \text{subject to} \quad & w_i(k) = w_{i,\text{meas}}(k), \quad i = 1, \ldots, d, \quad k \in \mathcal{C}_i, \end{aligned} \tag{4.18}$$

where the clean data is treated as exact, and the corrupted data is corrected.

4.3.6 System identification

The key idea of (4.4) is to find signal estimates for the measurements $w_{\text{meas}}(k)$ that can be modeled by a low-order dynamical system. However, the low-order system that the estimate $w(k)$ corresponds to is never explicitly computed. This is generally beneficial, as the details of the model do not need to be considered in order to accomplish the signal processing task. If it is however useful to compute the underlying system, this can done using subspace methods.

This section will demonstrate how to compute systems in their state space representation (introduced in Section 3.1) because of their popularity and usefulness. So, the aim of this task is to find estimates for the system matrices (A, B, C, D) and the initial state $x(0)$. The system identification process can be summarized by the following steps, adapted from the MOESP method [VV07, VVB07]:

1. Partition the signals into inputs $u(k)$ and outputs $y(k)$ as

$$w(k) = \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}. \tag{4.19}$$

2. Solve any of the above methods, for example (4.4), while replacing $W_{r,N}$ with

$$F(u, y) = \begin{bmatrix} U_{r,N} \\ Y_{r,N} \end{bmatrix}. \tag{4.20}$$

   Because $W_{r,N}$ and $F(u, y)$ are row permutations of each other, they have the same rank and trace norm, that is, $\|W_{r,N}\|_* = \|F(u, y)\|_*$.

3. Compute the LQ factorization

$$\begin{bmatrix} U_{r,N} \\ Y_{r,N} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \tag{4.21}$$

such that $Q_1^\top Q_1 = I$, $Q_2^\top Q_2 = I$, and $Q_1^\top Q_2 = 0$.

4. Compute the SVD

$$L_{22} = \begin{bmatrix} \bar{U}_1 & \bar{U}_2 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{V}_1^\top \\ \bar{V}_2^\top \end{bmatrix}. \tag{4.22}$$

5. Noting that $\text{range}(L_{22}) = \text{range}(\mathcal{O}_r)$, estimate matrices A and C from

$$\bar{U}_1 = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix}. \tag{4.23}$$

Van Overschee [VD12] provides efficient algorithms to do this using least squares. It is possible to constrain A to have eigenvalues $|\lambda_i(A)| < 1$, so that the system is stable.

6. Now that the input–output equation (3.5) is linear in its unknowns, estimate B, D, and $x(0)$ using least squares minimization. For example,

$$(B, D, x(0)) = \underset{\hat{B},\hat{D},\hat{x}(0)}{\text{argmin}} \sum_{k=0}^{N-1} \|y_{\text{meas}}(k) - y_{\text{model}}(k)\|_2^2, \tag{4.24}$$

where

$$y_{\text{model}}(k) = CA^k\hat{x}(0) - \hat{D}u(k) - \left[\sum_{j=0}^{k-1} CA^{k-1-j}\hat{B}u(j)\right]. \tag{4.25}$$

4.3.7 Algorithms

The data methods presented in this chapter are posed as convex optimization problems that do not have closed-form solutions; they are instead solved using iterative algorithms. Importantly, the problems have a special structure by which they can be solved efficiently using a class of optimization algorithms called *proximal algorithms*. Chapter 7 reviews the theory of these algorithms. Section 7.4 provides the specific steps to apply these algorithms to the problems of this chapter.

4.4 Simulations

4.4.1 Missing data

Imputation quality was measured by removing CBFV sections from measured data, imputing an estimate by solving (4.6), and comparing this estimate to the measured section that was removed. This process was repeated over 169 patient episodes for missing sections of 2 s in length. The element-wise median [IQR] absolute error for the CBFV model was 4.77 [2.00–10.38] cm/s (5.87%), and for the ICP model, 4.00 [1.68–8.32] cm/s (4.87%).

Example reconstructions using the CBFV model are shown in Fig. 4.4. Because this model only requires ABP and CBFV, these results can be achieved without knowledge of ICP. The same examples are also reconstructed using the ICP model in Fig. 4.5. The results show improvement, which is expected given that there is additional useful data available to the model. Imputation quality degrades gracefully as the length of missing data increases. This relationship is shown for the CBFV model in Fig. 4.6.

4.4.2 Artifact rejection

Artifacts in the 169 patients were removed by solving (4.4). The method is capable of removing artifactual sections of data without degrading the quality of clean sections. Examples of artifact rejection outcomes for different scenarios using the CBFV model are shown in Figs. 4.7, 4.8, and 4.9.

Because the true uncorrupted CBFV signal is not known, it is difficult to quantify the results of the algorithm, although visual results are satisfactory. For the same reason, it is also difficult to cross-validate an 'optimal' value of λ. The effect of varying λ is demonstrated in Fig. 4.10; higher values more closely resemble measurements, while lower values aggressively remove higher-order dynamics from the signal. 'Optimal' values of λ depend on signal length, because the two terms in (4.4) do not scale similarly as N increases. Otherwise, visually tuning a value of λ is a brief task.

The advantage of knowing prior location of artifacts is demonstrated in Fig. 4.11. By

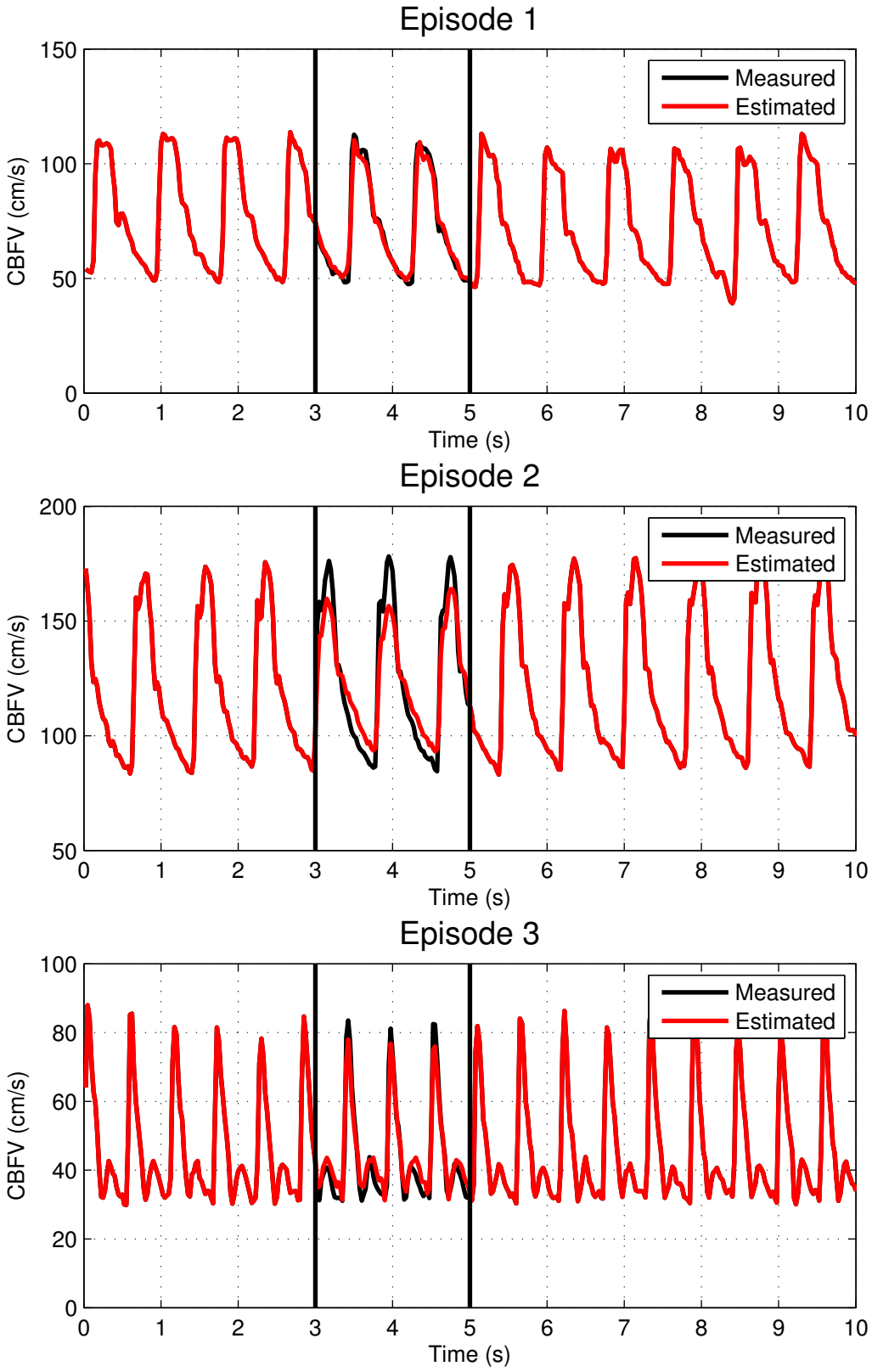

Figure 4.4: Imputation of missing CBFV using the CBFV model. Data was removed between 3–5 s. These results can be achieved without knowing patient ICP.

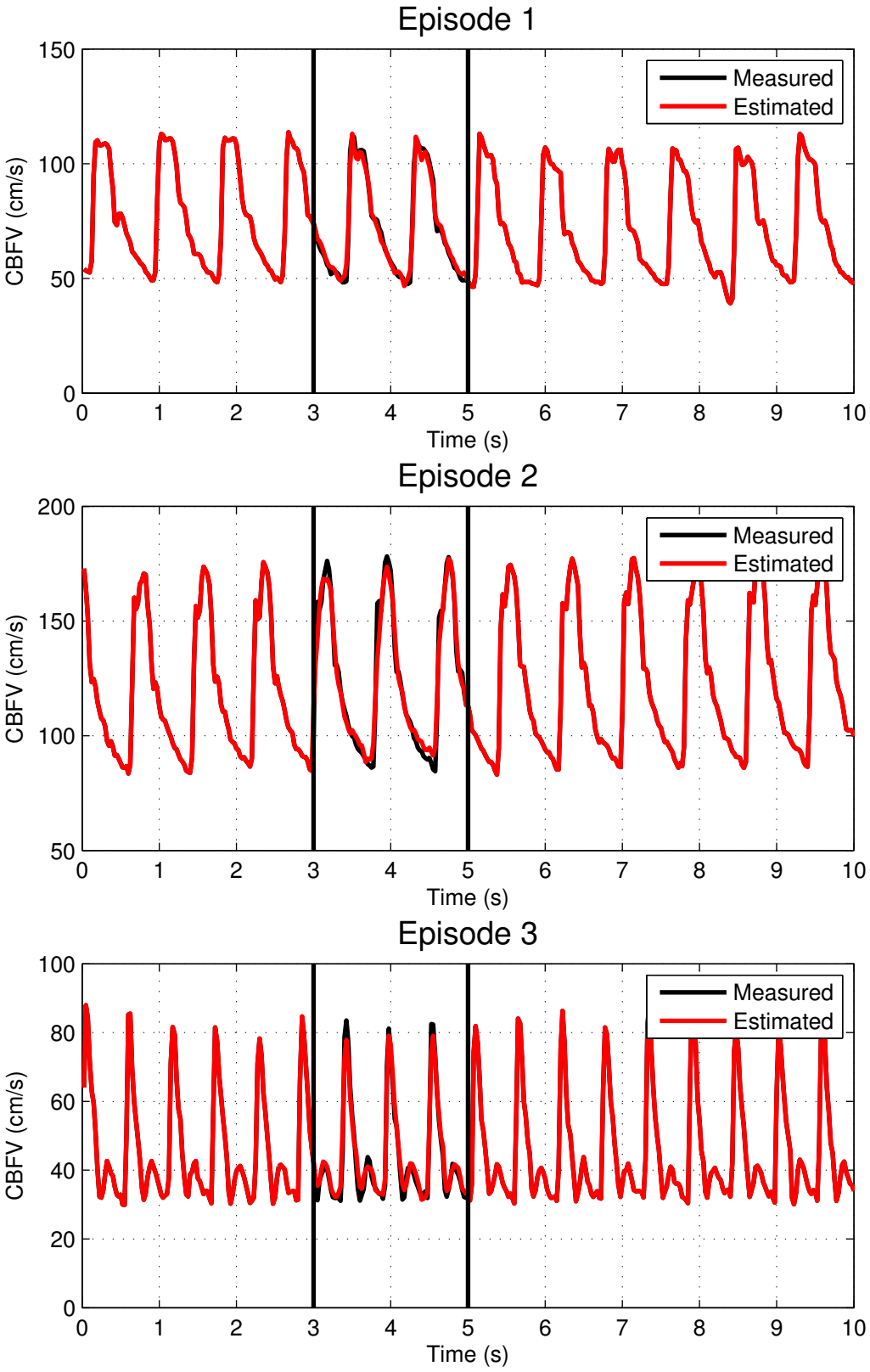

Figure 4.5: Imputation of missing CBFV using the ICP model. Data was removed between 3–5 s. These results require knowledge of patient ICP.

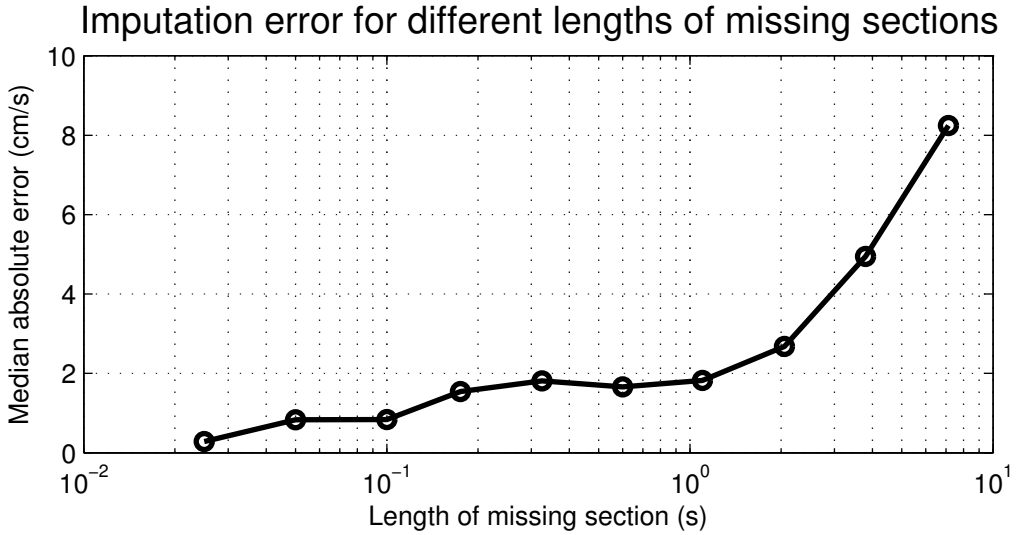

Figure 4.6: Median absolute imputation error by the CBFV and ICP models for different lengths of missing data.

visually examining the CBFV signal, it is apparent that an artifact occurs in the section $t \in (4.5, 6.5)$ s. For this section, $\lambda_i^{\text{corrupt}} = 0.001$. The problem is solved using (4.18). Because the artifact is aggressively adjusted while preserving the waveform in the clean section, this approach provides the best results. However, it requires users to define the section of corrupted data in a priori.

4.4.3 Structured artifact rejection

An example of structure artifact removal is given in Fig. 4.12 and Fig. 4.13 using a block length of $q = 20$ (0.5 s), and $\lambda = 0.35$. The method decomposes the measured CBFV signal into the sum of a CBFV estimate, generated from a low-order dynamical system, and a small set of nonzero elementary artifact terms of length 0.5 s. The result produces six nonzero residual terms, demonstrating the expected block-sparse correction structure.

Although the output of this method has a useful interpretation, in that each of the artifacts can be retrieved as individual objects (see Fig. 4.13), there was no evidence that this approach improved upon artifact rejection from the ℓ_1-norm approach of (4.4). Additionally, the introduction of an additional parameter to tune, q, suggests that this method is more

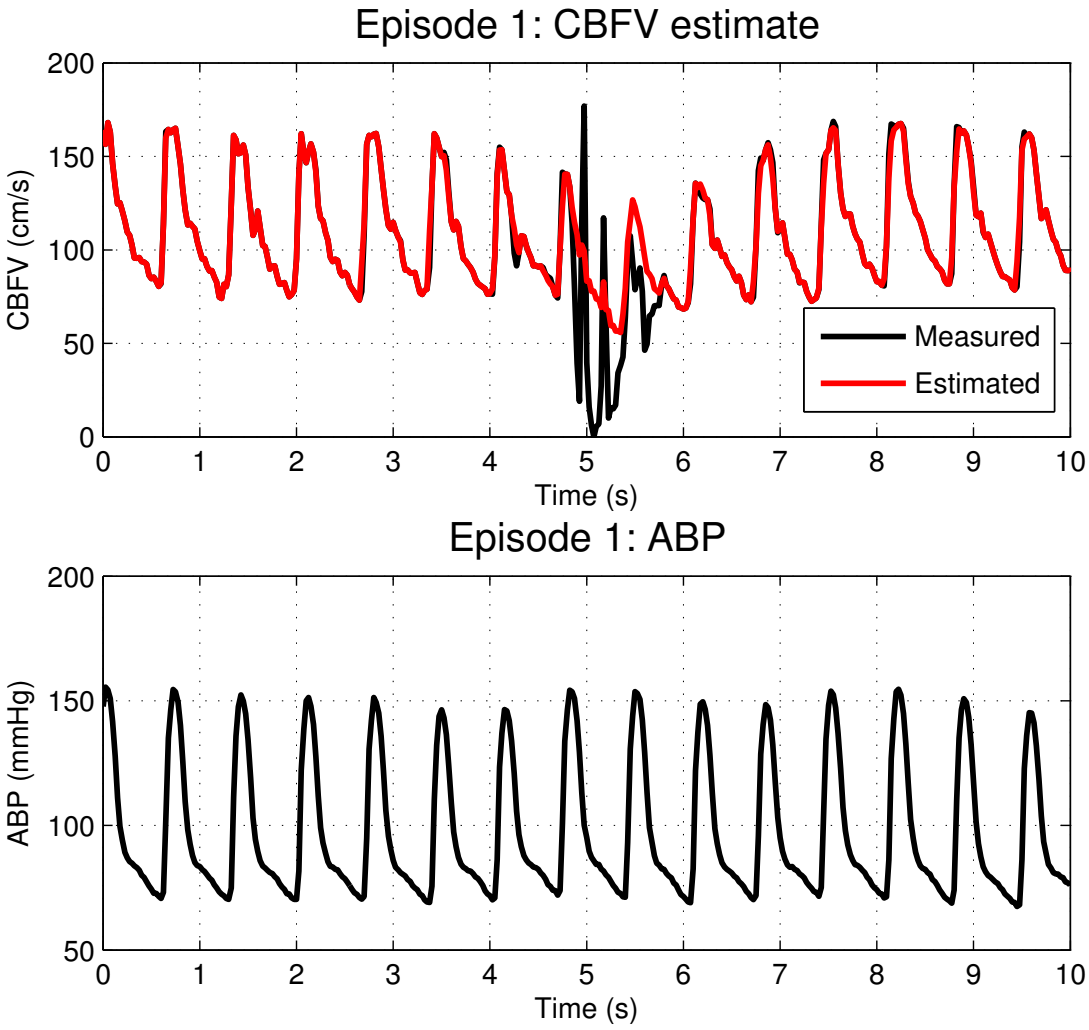

Figure 4.7: Removal of a large artifact using the CBFV model. Only ABP and CBFV are required to achieve this quality of rejection.

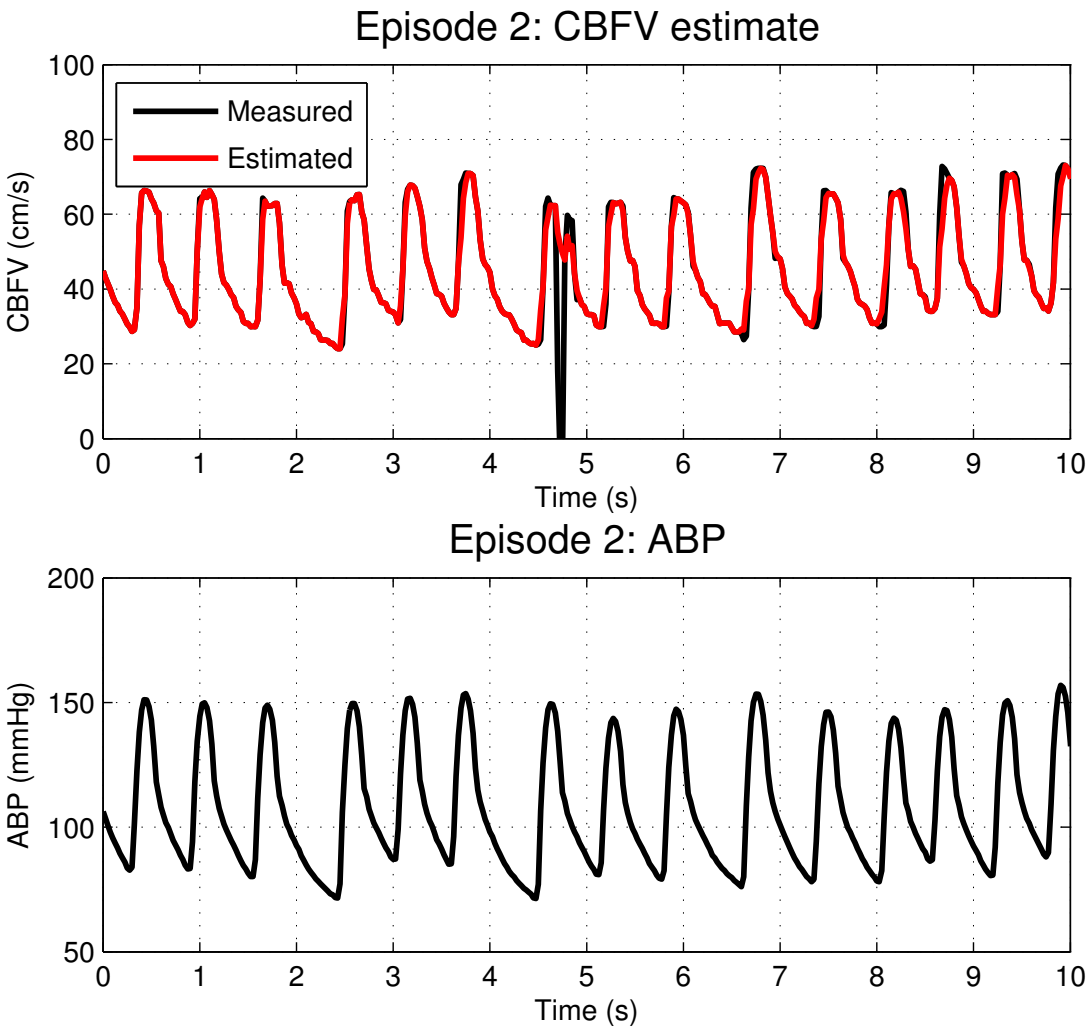

Figure 4.8: Removal of a sparse, sharp artifact using the CBFV model.

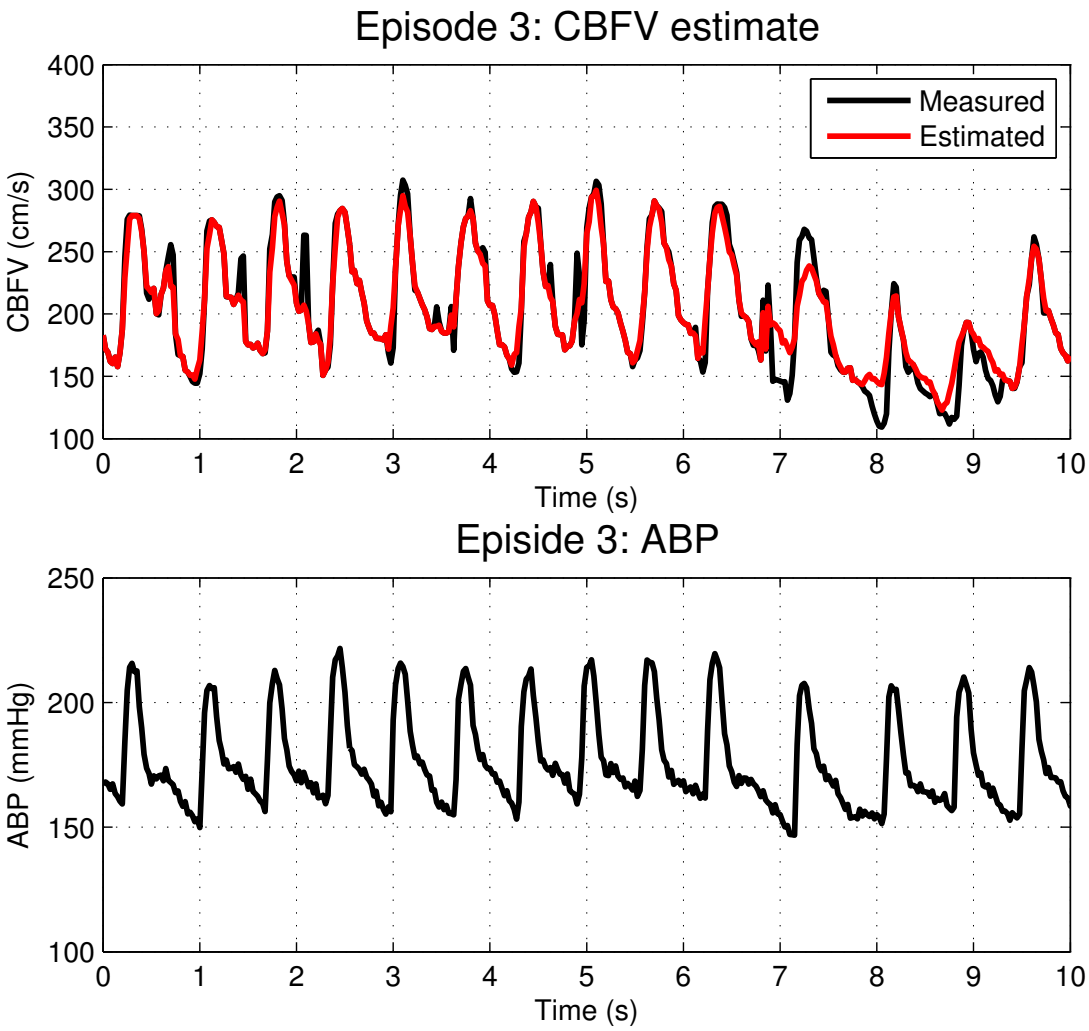

Figure 4.9: Removal of frequent disturbances using the CBFV model.

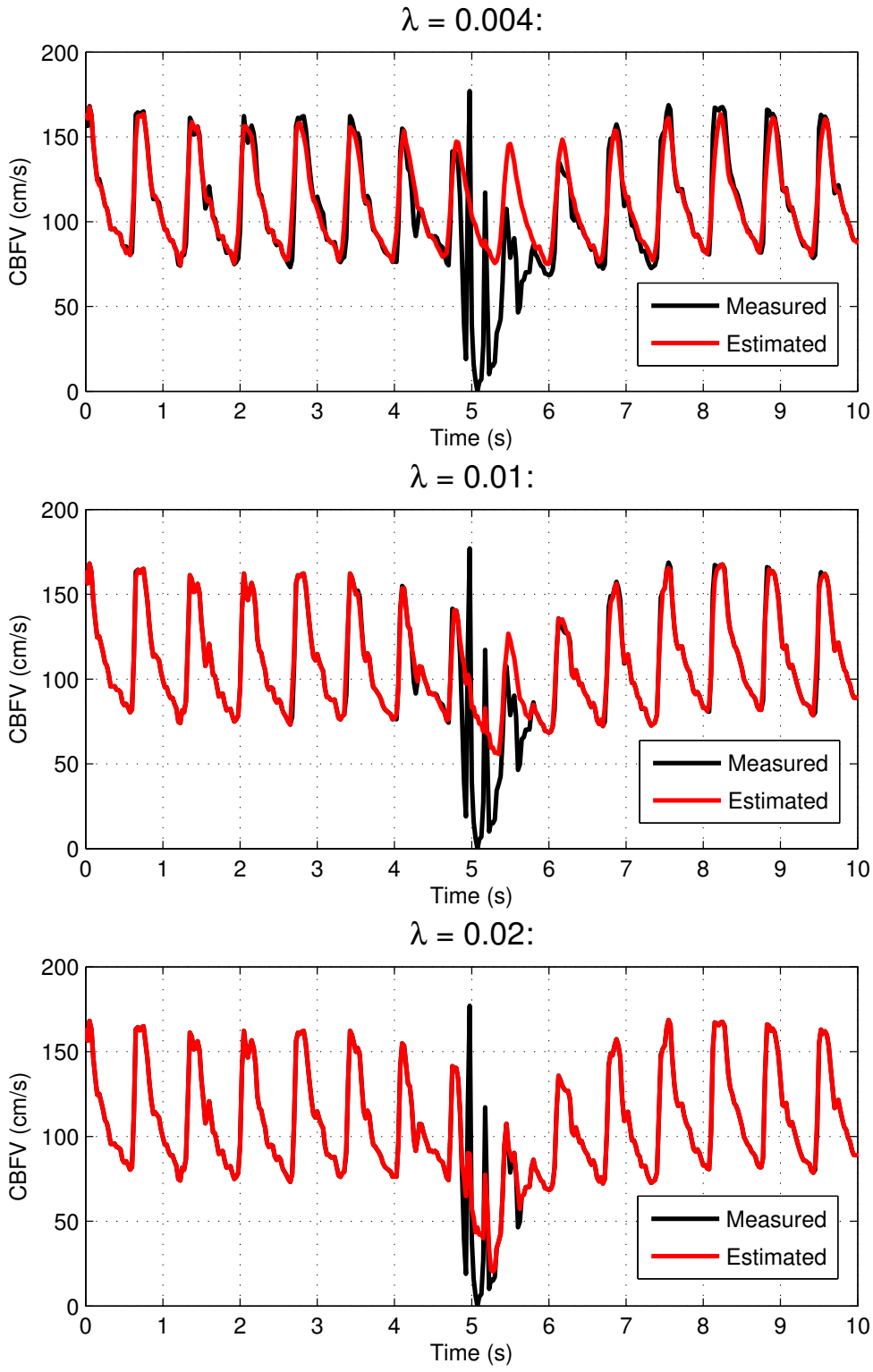

Figure 4.10: Artifact rejection for varying values of λ, using the CBFV model. Lower values of λ produce signals with lower-order dynamics, while higher values more closely resemble the measured data.

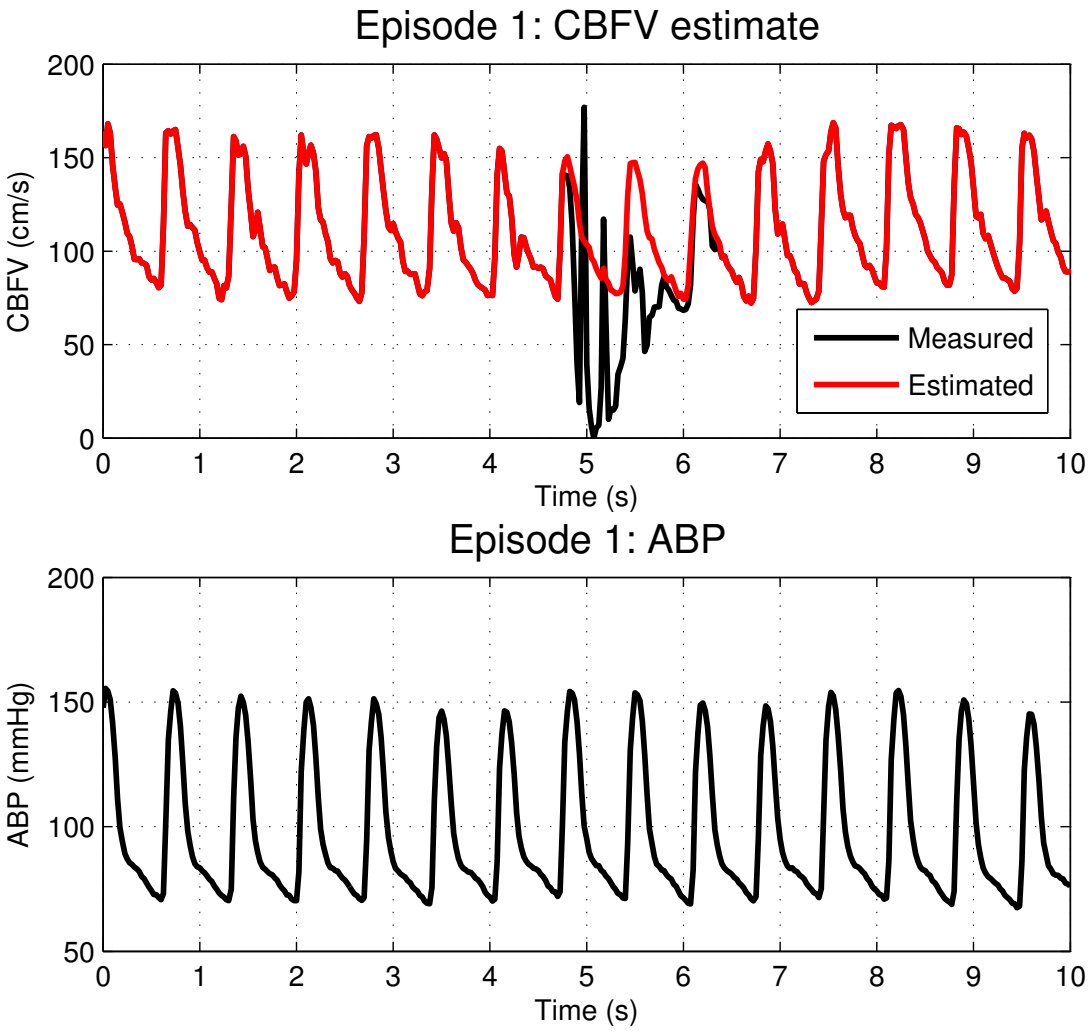

Figure 4.11: Outcomes with prior knowledge that an artifact occurs in the range $t \in (4.5, 6.5)$ s, using the CBFV model. The clean data is kept exact, and the corrupted section is optimized over.

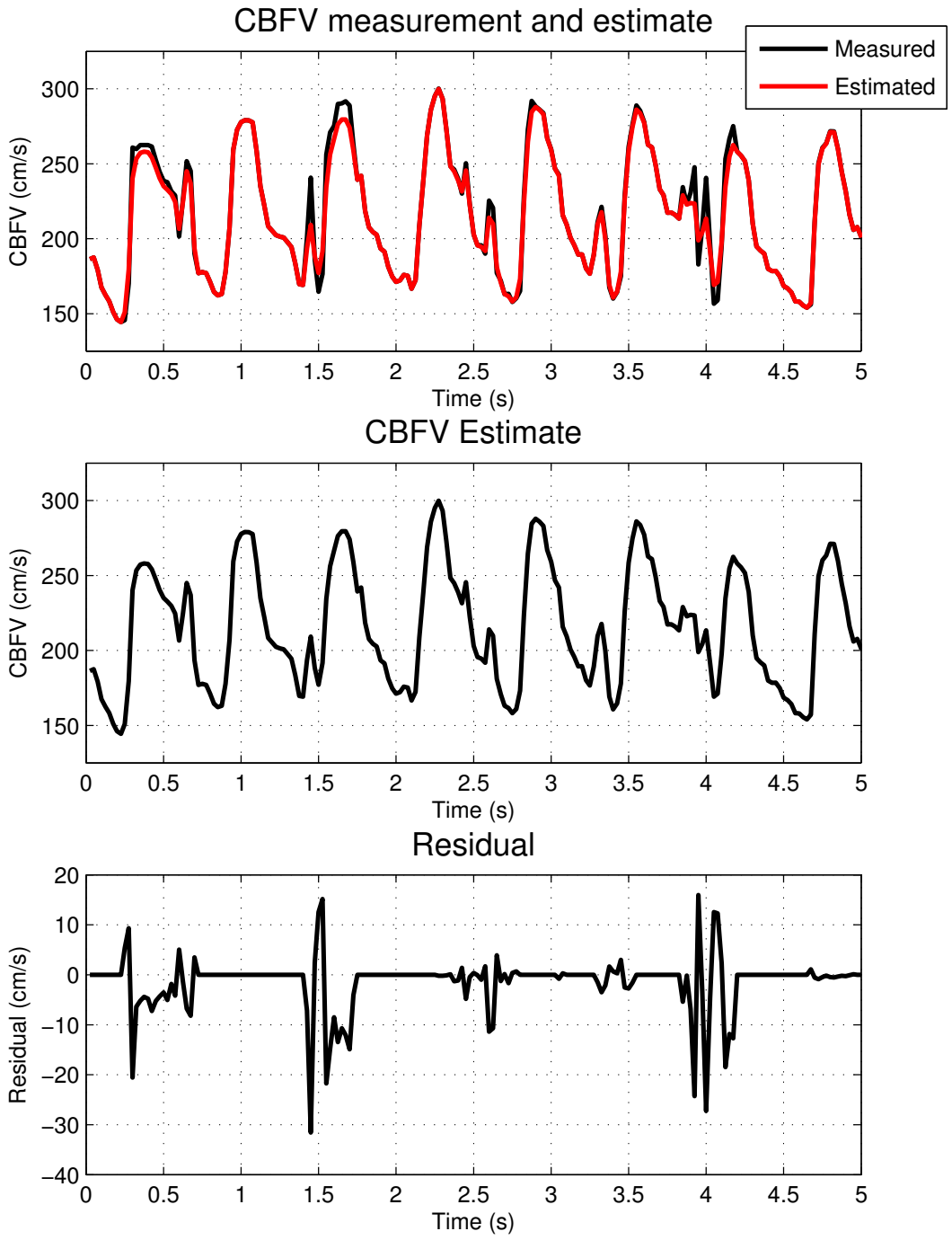

Figure 4.12: Overview of CBFV structured artifact removal via (4.12), using $q = 20$ and $\lambda = 0.35$.

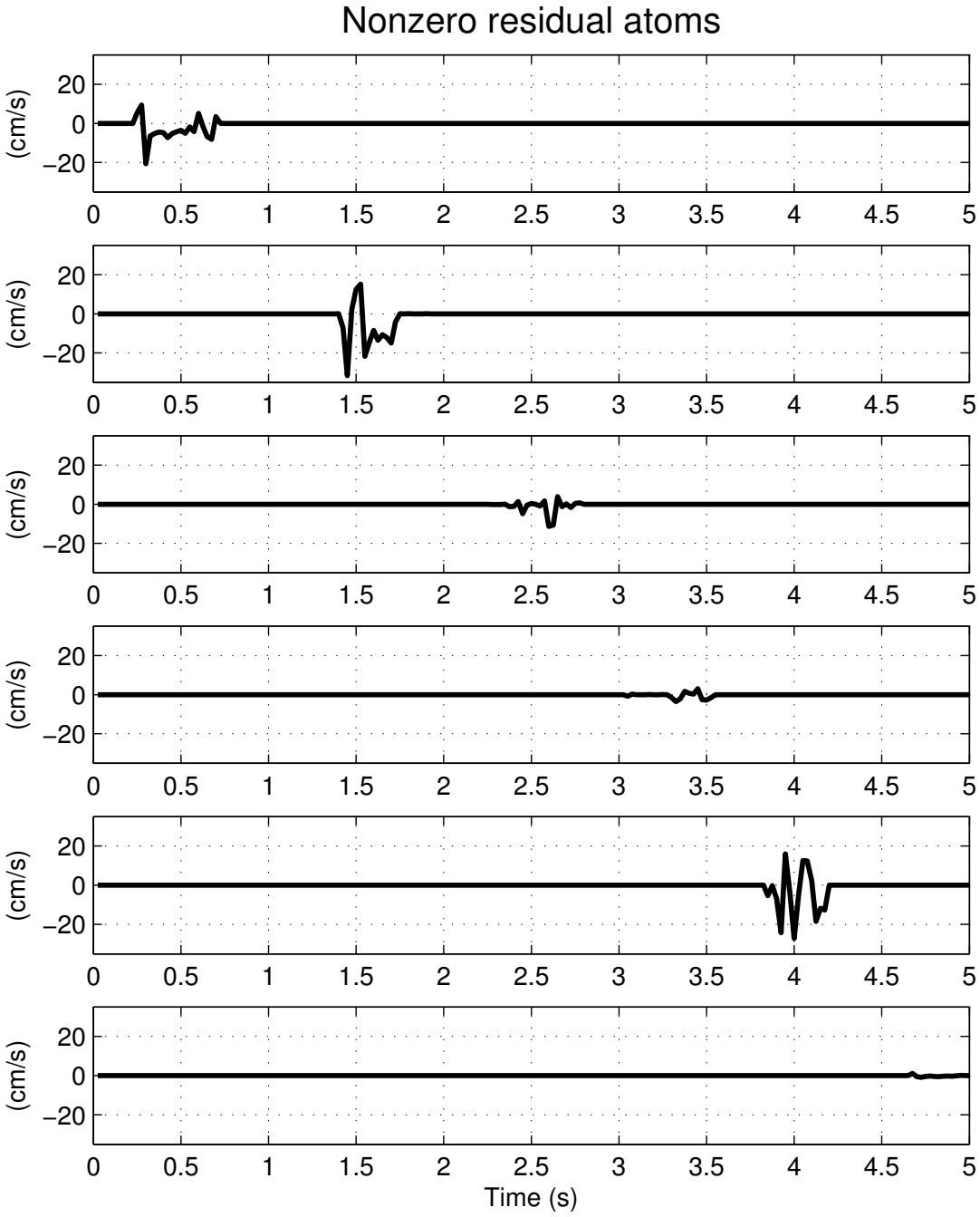

Figure 4.13: Structure of nonzero residual *atoms* produced by the CBFV structural artifact rejection method (4.12). These six nonzero residuals start at $t = 0.25, 1.43, 2.28, 3.13, 3.85, 4.68$ s and are 0.5 s in length.

difficult to quickly adopt for new use cases.

4.5 Discussion

This chapter presented a suite of convex optimization methods to reject artifacts and impute missing data from CBFV signals. The key principle of these methods was to impose the prior assumption that the signals were well-approximated by a low-order dynamical model. Because of the flexibility of the convex optimization framework, the approach was easily extended to a variety of scenarios. Simulations performed the artifact rejection and missing data imputation tasks well on clinical data.

As a standalone method, future steps will focus on validation using a large-scale data set. The approaches should be tested on a large body of patient data to confirm its robustness to different clinical scenarios. Possible research questions include determining whether additional physiological signals improve outcomes; whether differences in diagnosis (TBI, aSAH, etc.) affect quality of signal enhancement; determining clinically-optimal values of λ, and the conditions in which it varies; suggesting a suitable signal length N over which to perform operations; and, for structured artifact rejection, a clinically-optimal value of q. While these questions must be investigated before the approach is relevant at point-of-care, it is immediately useful to researchers working with physiological signals.

In the context of this dissertation, the takeaways of this chapter are that **(a)** ICP can be accurately modeled as a low-order LTI system driven by ABP and CBFV, and **(b)** the regularization penalty $\|W_{r,N}\|_*$ can accurately impute missing data in this relationship by finding values that imply a low-order system.

CHAPTER 5

Methods to cluster patient models as dynamical systems

Note: An earlier version of this chapter was presented at the IEEE Biomedical and Health Informatics Conference, 2018, as *Clustering Patient Models of Intracranial Pressure Dynamics* (Gunn, C.A., Hu, X., Vandenberghe, L.).

5.1 Introduction

Clustering is one of the principal tasks in machine learning, in which objects are categorized into groups by a similarity or distance metric. This idea has been applied to a variety of objects: clustering numerical data is a classical topic [DHS12], and kernel methods have extended early approaches to more complex objects, such as strings of text, and images [SC04]. Less attention has however been given to clustering models of *dynamical systems*, such as LTI models.

In the setting of biomedical signal processing, it is common to work with dictionaries of patient-specific LTI models. These models typically share the same structure—such as model representation and signal channels—but have unique parameterizations. Clustering these patient-specific models by their dynamics therefore has the potential to group patients, past or present, by a new metric that may reveal unique information on diagnosis or progress.

NICP estimation provides a key motivation for such a method. Chapter 4 demonstrated that patient-specific models can accurately predict patient ICP using only noninvasive signals, but it is not possible to fit these models without prior ICP information. In contrast,

models fit across the entire population performed poorly due to large inter-patient variability. Interestingly, it has been noted from clinical anecdote that patients' ICP dynamics may be classified into relatively few patterns [CP04]. This observation suggests that it may be possible to cluster such patients using their available LTI dynamics so that more patient-targeted ICP models can be constructed, improving ICP estimation without the need for any invasive measurements.

There are many other biomedical examples of patient-specific LTI models, typically autoregressive-moving-average (ARMA) models, where such a clustering method could be applied to patient dictionaries. Examples of applications to existing models include MRI time series data [FHP95], electrocardiogram (ECG) analysis [Ste02], blood glucose measurement [LGB10] and hyperglycemia detection [DNZ13], and respiratory monitoring [LC11]. This approach is also relevant to numerous applications beyond physiology. In finance, ARMA models of a publicly-traded company's financial signals can be clustered to find stocks with similar dynamics [LL95]. In economics, models can be constructed around a country's microeconomic signals, to determine how similar two countries' response would be to unprecedented economic events [KT02]. In biology, a wide array of time series models can be used—from gait [LBP09] to population growth [FSP98]—to compare the dynamics of species or colonies.

This chapter introduces a method to cluster LTI models by the similarity of their dynamics. The method is based on an extension of the *sparse subspace clustering* proposed method by Elhamifar & Vidal [EV09], and clusters models by their observability matrices. The approach is validated and tested on models of ICP dynamics, and shown in a short example to add value to an NICP estimator. Finally, several variations of the method are proposed and tested.

5.2 Background

5.2.1 Data model

This chapter considers a dictionary of L LTI models $\Sigma_1, \ldots, \Sigma_L$. Each models takes the form

$$\Sigma_i : \quad \begin{cases} x(k+1) &= A_i x(k) + K_i e(k) \\ y(k) &= C_i x(k), \end{cases} . \tag{5.1}$$

where $y(k) \in \mathbb{R}^p$ contains the p observed signals, $x(k) \in \mathbb{R}^n$ is the unobserved *state* vector, and $e(k) \sim \mathcal{N}(0,1)$ is a noise perturbation. All models are assumed to have a uniform order n. The matrices (A_i, K_i, C_i) define the system parameters. The *extended observability matrix* of Σ_i is

$$\mathcal{O}_i = \begin{bmatrix} C_i \\ C_i A_i \\ C_i A_i^2 \\ \vdots \\ C_i A_i^{q-1} \end{bmatrix} \in \mathbb{R}^{pq \times n}. \tag{5.2}$$

This matrix has a useful interpretation: consider that in the noise-free case ($e(k) = 0$), with an initial state $x(0)$, the system trajectory is given by

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(q-1) \end{bmatrix} = \begin{bmatrix} C_i \\ C_i A_i \\ \vdots \\ C_i A_i^{q-1} \end{bmatrix} x(0) = \mathcal{O}_i x(0). \tag{5.3}$$

So, the column space of $\mathcal{O}_r$ is the subspace that possible trajectories lie in.

5.2.2 Metrics to compare dynamical systems

One of the challenges to clustering LTI models is that there is no widely-accepted metric for the 'distance' between two models. This section will review the established options. A small number have been proposed; an overview of them is given in [KB14]. Two interesting choices are the Martin distance [Mar00, DD02], and Cauchy–Binet kernels [VSV07]. Both of these

methods operate on ARMA models, and measure generalized notions of output trajectories to quantify model similarity.

Martin distance. In its original interpretation, the Martin distance [Mar00] compares the *cepstra* of two models. Martin notes that the metric has several consistent properties that suggest that it is a 'natural' metric for comparing these systems. One interpretation is that the method computes a filter to 'whiten' the spectrum of the first model, and then applies this filter to the second model. If the models are similar, the second model's spectrum should also be nearly whitened. The Martin distance measures this deviation from a white spectrum.

De Kock & De Moor [DD00, DD02] demonstrated that this metric has another important interpretation that is related to the principal angles between subspaces. Given two systems Σ_i and Σ_j with extended observability matrices $\mathcal{O}_i$ and $\mathcal{O}_j$ respectively, the subspaces of which forming the principal angles $\theta_1, \ldots, \theta_p$, then the Martin distance can be expressed as

$$d_M(\Sigma_1, \Sigma_2) = \sqrt{\ln \prod_{i=1}^{p} \frac{1}{\cos^2 \theta_i}}. \tag{5.4}$$

This interpretation is interesting because it demonstrates that the column spaces of the observability matrices of two systems can be used as a principled metric for their comparison. As noted above, these matrices represent generalized trajectories. Hence, the Martin distance can also be thought of as measuring the difference in trajectories between systems subjected to arbitrary initial conditions.

Cauchy–Binet kernels. Another related metric of system similarity is the family of Cauchy–Binet kernels [VSV07]. These kernels model the inner product of two ARMA systems as the inner product of their trajectories. There are several variations; one example is the *trace kernel*, which compares (Σ_i, Σ_j) as a converging weighted inner product of their outputs $y_i(k)$ and $y_j(k)$ over an infinite horizon:

$$d_{\mathrm{BC}}(\Sigma_1, \Sigma_2) = \mathbb{E}_{e(k)} \left[\sum_{k=0}^{\infty} e^{-\lambda k} y_i(k)^\top W y_j(k) \right], \tag{5.5}$$

where $W > 0$ is a user-defined weight matrix. These kernels share a similar motivation to the Martin distance and have some interesting theoretical properties, but they are less easily applied to datasets in practice.

5.2.3 Background on system clustering and classification

A framework for clustering and classifying ARMA models has been developed by Ravichandran et al. [RCV09, RCV13] for the problem of dynamic texture recognition. In this problem, videos of textures, such as smoke, water, or fire, are modeled by ARMA processes. The authors integrate the Martin distance with a sampling-based mapping function to construct features that lie in a Euclidean space. These features could then be clustered using classical centroid-based techniques, such as k-means clustering [Har75, HW79].

5.2.4 Outline of new approach

This chapter will introduce a new framework for clustering LTI models. As with the Martin distance, this new metric will compare observability matrices of systems—specifically, the column spaces of these matrices. The clustering step of this framework is a novel extension to a previously reported method for clustering vector subspaces.

Additionally, a new metric is derived from this approach which will be called the *subspace reconstruction distance* between two matrices. This metric will be compared to the Martin distance and used in a baseline clustering method against which to compare results.

5.3 Method development

5.3.1 Review: Clustering vectors by subspace

Elhamifar & Vidal [EV09, EV13] proposed a method to cluster vector-valued elements that lie in a union of low-dimensional subspaces. The method is called *sparse subspace clustering* (SSC). Multiple extensions have been proposed in literature [CMS14, SEC14, SC12].

The SSC approach exploits the *self-expressiveness* property of data, which states that elements that lie in the same subspace can be expressed as linear combinations of one another. The method uses this principle to find a sparse and unique representation of each element as a linear combination of other elements, revealing which elements share subspaces. Because finding the *most-sparse* reconstruction, equivalent to ℓ_0-norm minimization, is a hard problem, the method uses an ℓ_1-norm convex relaxation.

Given L vector-valued data objects $y(i) \in \mathbb{R}^n$, for $i = 1, \ldots, L$, SSC seeks to assign each of the L objects into $k < L$ clusters. To do this, first the data is arranged in the matrix form

$$Y = \begin{bmatrix} y(1) & \cdots & y(L) \end{bmatrix} \in \mathbb{R}^{n \times L}. \tag{5.6}$$

For each $y(i)$, a *weight vector* $g_i \in \mathbb{R}^L$ is introduced such that

$$y(i) = Y g_i, \tag{5.7}$$

which specifies that $y(i)$ is a linear combination of the set of vectors $y(1), \ldots, y(L)$. Additionally, $y(i)$ should not be used to reconstruct itself, so $g_{ii} = 0$. The following optimization problem then seeks a sparse representation of g_i given these constraints:

$$\begin{aligned} \text{minimize} \quad & \|g_i\|_1 \\ \text{subject to} \quad & y(i) = Y g_i \\ & g_{ii} = 0. \end{aligned} \tag{5.8}$$

When repeatedly solved for $i = 1, \ldots, L$, the problem can be written in matrix form:

$$\begin{aligned} \text{minimize} \quad & \|G\|_{11} \\ \text{subject to} \quad & Y = YG \\ & \operatorname{diag}(G) = 0, \end{aligned} \tag{5.9}$$

where $\|\cdot\|_{11}$ is the sum of the absolute values of every matrix entry, and

$$G = \begin{bmatrix} g_{11} & \cdots & g_{L1} \\ \vdots & \ddots & \vdots \\ g_{1L} & \cdots & g_{LL} \end{bmatrix} \in \mathbb{R}^{L \times L}. \tag{5.10}$$

The authors provide details on how to solve these optimization problems using the ADMM algorithm [EV13].

When data is corrupted by noise or artifacts, it is unlikely to fall exactly in a small set of subspaces. In this case, the reconstruction equation $Y = YG$ can be approximated. A general form for this is

$$\begin{aligned} \text{minimize} \quad & \|G\|_{11} + \gamma\|E\|_{11} + \frac{\lambda}{2}\|Z\|_F^2 \\ \text{subject to} \quad & Y = YG + E + Z \\ & \operatorname{diag}(G) = 0, \end{aligned} \tag{5.11}$$

where E and Z are residual or correction variables to solve, and $(\gamma, \lambda) > 0$ are constants. This combination of penalties using the ℓ_{11}-norm and Frobenius norm is similar to elastic net regularization [ZH05]. When the data is only corrupted by noise, (5.11) can be simplified to the convenient form

$$\begin{aligned} \text{minimize} \quad & \|G\|_{11} + \frac{\lambda}{2}\|Y - YG\|_F^2 \\ \text{subject to} \quad & \operatorname{diag}(G) = 0. \end{aligned} \tag{5.12}$$

The value of λ can be loosely interpreted as a tolerance for correcting the data from noise corruption; lower values will allow greater deviations from $Y = YG$ in order to promote sparsity in G.

The remaining steps of the SSC method are reasonably simple. First, a symmetric weight matrix $V = |G| + |G|^\top \in \mathbb{S}^{L \times L}$ is constructed to remove the directionality of the weights in G. This choice of V can be interpreted as a sparse proximity graph between the elements $y(1), \ldots, y(L)$. Then, a spectral clustering method is applied to V to identify k distinct clusters in the data. These methods often require a user-defined choice of k.

Spectral clustering algorithms are used to cluster nodes that are connected over a graph. In the SSC method, each data element $y(i)$ is a node on a graph, and the weight $v_{ij} > 0$ is the edge weight between nodes $y(i)$ and $y(j)$. Spectral clustering algorithms attempt to find a *partition* of this graph that will cut lesser-weighted edges while preserving greater-weighted edges [Von07]. The final number of clusters k is defined by the user.

The general spectral clustering procedure is concisely summarized by Von Luxburg in [Von07]. Two common variations of the method are Shi & Malik clustering [SM00], and Ng, Jordan and Weiss clustering [NJW02]. The methods differ essentially by unique choices of the *Laplacian matrix* of the graph. This dissertation will use Shi & Malik clustering.

5.3.2 Clustering matrices by column space

The first contribution of this chapter is to extend the SSC method to the matrix case. The method is called *sparse column space clustering*. In this scenario, the L data elements are matrices $Y(i) \in \mathbb{R}^{m\times n}$ whose column spaces lie in the union of p subspaces. The aim of this method is to again assign each of the L objects into $k < L$ clusters. This extension will be achieved by using a *group sparsity* approach.

First, the data matrix is redefined as

$$\hat{Y} = \begin{bmatrix} Y(1) & \cdots & Y(L) \end{bmatrix} \in \mathbb{R}^{m\times nL}, \tag{5.13}$$

and the weight matrix as

$$G = \begin{bmatrix} G_{11} & \cdots & G_{L1} \\ \vdots & \ddots & \vdots \\ G_{1L} & \cdots & G_{LL} \end{bmatrix} \in \mathbb{R}^{nL\times nL}. \tag{5.14}$$

Here, each $G_{ij} \in \mathbb{R}^{n\times n}$ is a submatrix defining the weights to reconstruct $Y(i)$ as a linear combination of each $Y(j)$. At a per column level, each element $[G_{ij}]_{k\ell}$ is the weight to reconstruct the column $[Y(i)]_k$ as a linear combination of each column $[Y(j)]_\ell$. The first step of the SSC problem can then be rewritten with a group sparsity approach as

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{L}\sum_{j=1}^{L} \|G_{ij}\|_F \\ \text{subject to} \quad & \hat{Y} = \hat{Y}G \\ & G_{ii} = 0, \quad i = 1, \ldots, L. \end{aligned} \tag{5.15}$$

As in (5.12), data that is corrupted by noise may not lie exactly in a union of subspaces. In

these cases, the reconstruction equation $\hat{Y} = \hat{Y}G$ is approximated, giving

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{L}\sum_{j=1}^{L} \|G_{ij}\|_F + \frac{\lambda}{2}\|\hat{Y} - \hat{Y}G\|_F^2 \\ \text{subject to} \quad & G_{ii} = 0, \quad i = 1, \ldots, L, \end{aligned} \tag{5.16}$$

for $\lambda > 0$.

The sums of Frobenius norms $\|\cdot\|_F$ in (5.16) are the matrix equivalent of the ℓ_1/ℓ_2-norm used on vector data to promote group sparsity. This penalty promotes a sparsity pattern between each matrix G_{ij}, but not within the individual entries of G_{ij} itself. Other choices of norm penalty are possible; the effect of this choice is investigated in Section 5.5.2. An algorithm to solve problems of the form of (5.16) will be outlined in Section 5.3.4.

In this scenario, the symmetric matrix $|G| + |G|^\top$ provides the bidirectional edge weights between every pair of *columns* in the data. To compute edge weights between pairs of *matrices* $(Y(i), Y(j))$ in the data, $V \in \mathbb{R}^{L\times L}$ is found by collapsing the group sparsity structure:

$$V_{ij} = \|G_{ij}\|_F^2 + \|G_{ji}\|_F^2. \tag{5.17}$$

In the final step, spectral clustering can be applied to this V, as in the original SSC method.

5.3.3 LTI model clustering method

Finally, to cluster LTI models by their dynamics, the sparse column space clustering method is applied to the observability matrices of the models. Hence, models are clustered by the column space of their observability mappings, as in [RCV13].

Consider the systems $\Sigma_1, \ldots, \Sigma_L$ that are of the same order n. Given their observability matrices $\mathcal{O}_1, \ldots, \mathcal{O}_L$, define

$$\mathcal{O} = \begin{bmatrix} \mathcal{O}_1 & \cdots & \mathcal{O}_L \end{bmatrix} \in \mathbb{R}^{pq\times nL}, \tag{5.18}$$

and apply the sparse column space clustering method (5.16) to the data. The problem then

becomes

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{L}\sum_{j=1}^{L} \|G_{ij}\|_F + \frac{\lambda}{2}\|\mathcal{O} - \mathcal{O}G\|_F^2 \\ \text{subject to} \quad & G_{ii} = 0, \quad i = 1, \ldots, L. \end{aligned} \tag{5.19}$$

The resulting steps to compute the model clusters are then summarized in Algorithm 1.

Algorithm 1 LTI model clustering

1: Given the models $\Sigma_1, \ldots, \Sigma_L$, compute the observability matrices $\mathcal{O}_1, \ldots, \mathcal{O}_L$ and the stacked observability matrix (5.18).
2: Select the constant $\lambda > 0$ and solve (5.19) to obtain G.
3: Compute the per-model weight matrix $V_{ij} = \|G_{ij}\|_F^2 + \|G_{ji}\|_F^2$.
4: Apply spectral clustering to V to retrieve the assignment of the models to k clusters.

5.3.4 Algorithm

The most expensive step of this new approach to system clustering is solving the convex optimization problem (5.19). The solution to this problem must be found using iterative convex optimization algorithms. Fortunately, (5.19) has special that can be exploited: it is *separable* (that is, it can be solved as multiple smaller problems in parallel), and it has a structure that can be solved efficiently using proximal algorithms. These algorithms are reviewed in Chapter 7. The algorithmic details to compute the solution to (5.19) are then given in Section 7.5.

5.4 Simulations

5.4.1 Clustering validation

The method is validated by testing it on a dataset with a known ground truth. Specifically, it is applied to a dictionary that contains multiple models per patient. The aim of the experiment is to determine whether the method clusters these same-patient models together.

The dictionary was constructed by randomly selecting 10 patients and dividing their time

series data into 10 episodes each, 36 heartbeats in length. This produced $L = 100$ episodes. For each episode, an LTI model of the form (5.1) was fit with the output signals

$$y(k) = \begin{bmatrix} \text{ABP}(k) \\ \text{CBFV}(k) \\ \text{ICP}(k) \end{bmatrix} \in \mathbb{R}^3, \tag{5.20}$$

and order $n = 3$ using the N4SID method [Lju07].

The 100 models were clustered into $k = 10$ groups. The method was found to be largely robust to choice of λ, so long as extreme values (i.e., extremely sparse matrices V) were avoided. Simulations used $\lambda = 1$, which produced a sparsity in V of 78%. The edge weights in V and the resulting cluster assignments are shown in Fig. 5.1. 92% of the models were correctly clustered. An example of how V represents edge weights in a graph, for a similar case with $L = 60$ models from 6 patients, is given in Fig. 5.2.

5.4.2 NICP estimation

The second experiment is a simple demonstration of how patient clustering adds value to NICP estimation. In the simulation, two NICP estimators are compared. The first is constructed from population-wide data. For the second, patients are first clustered, and cluster-specific estimators are constructed.

Both estimators predict a patient's mean ICP as an affine function of their mean CBFV, and were fit using robust linear regression (minimizing ℓ_1-norm error). The *baseline* estimator was trained and tested on two independent sets randomly drawn from the complete $L = 75$ patient dictionary. For the *clustered* estimator, patients were clustered a priori, and then estimators were constructed for each cluster. First, a patient-specific LTI model was fit to each of the patients using only the noninvasively-measured signals

$$y(k) = \begin{bmatrix} \text{ABP}(k) \\ \text{CBFV}(k) \end{bmatrix} \in \mathbb{R}^2, \tag{5.21}$$

and order $n = 3$. Then, system clustering was applied to the models to retrieve $k = 3$ clusters. Finally, estimators of the same form were trained and tested on independent sets

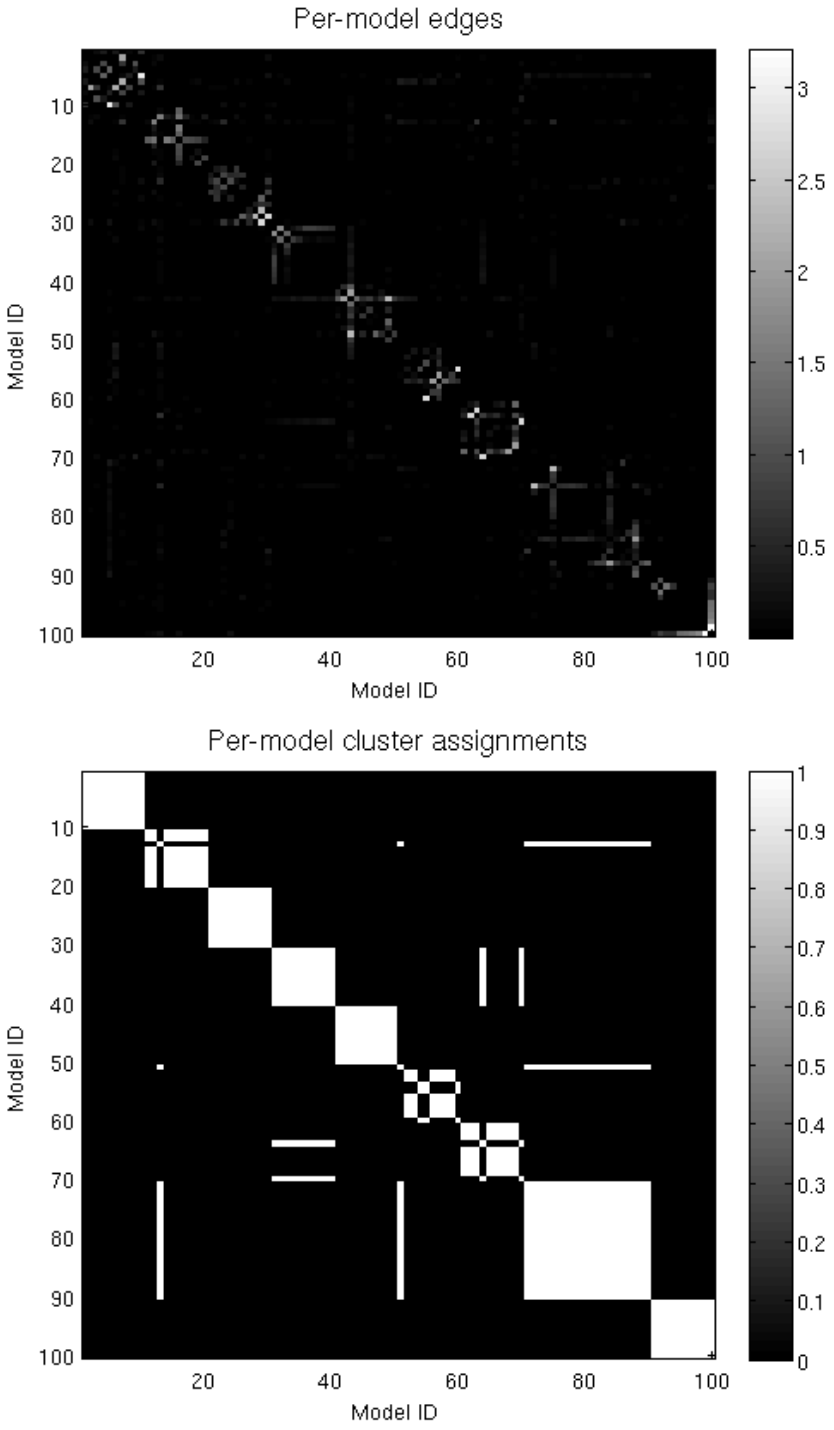

Figure 5.1: Edge weights between each model and corresponding cluster assignments. Models $1, \ldots, 10$ correspond to Patient 1, $11, \ldots, 20$ to Patient 2, etc.. 92% of the models were correctly clustered.

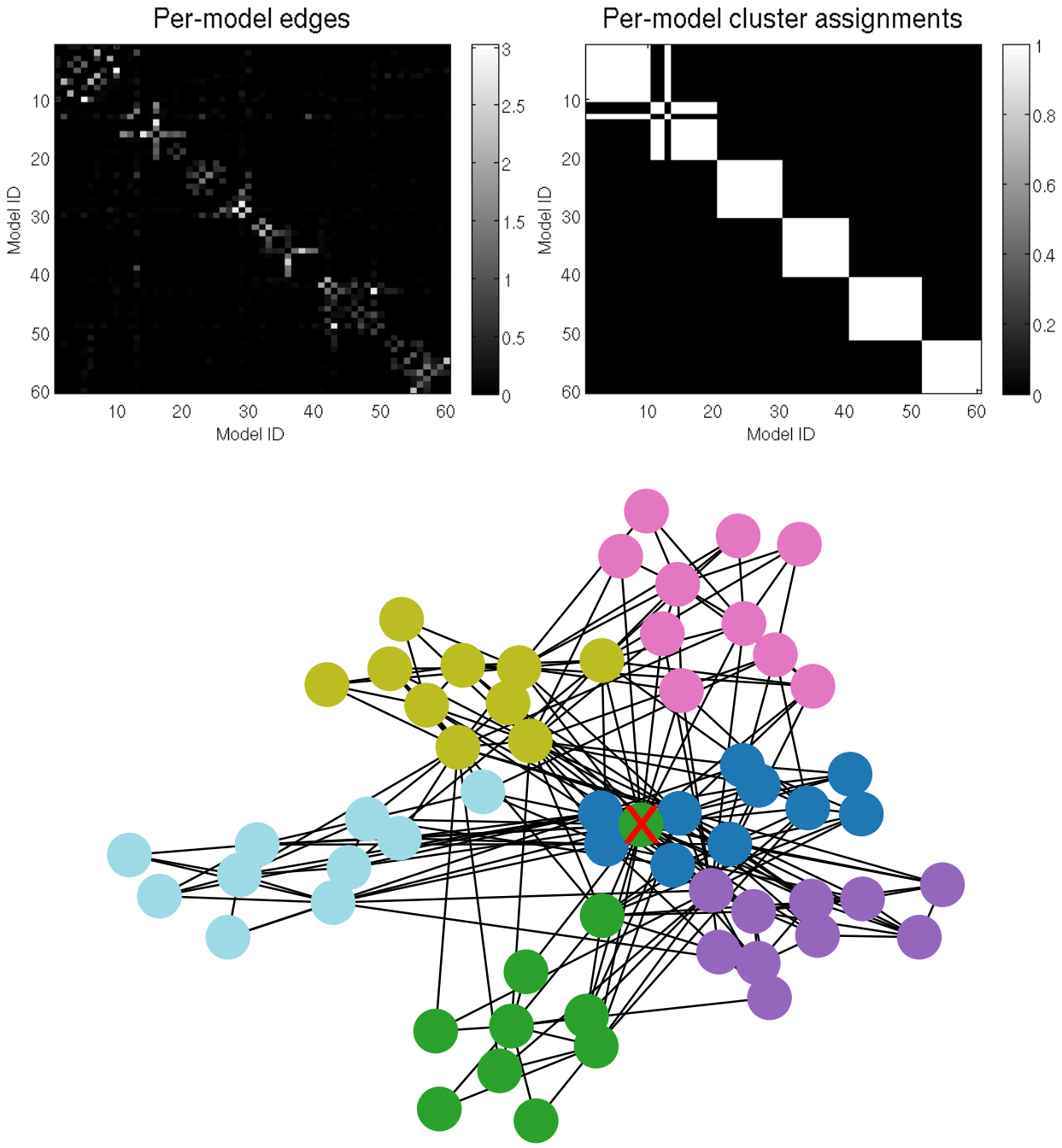

Figure 5.2: Adjacency of episodes in V via the Fruchterman–Reingold force–directed algorithm [FR91]. Nodes are colored by patient ID. The incorrectly-clustered nodes are marked by red crosses. Some lower-weight edges are omitted for clarity. This diagram was constructed using the NetworkX package [HSS08].

Table 5.1: NICP estimation error from two linear predictors. The first is trained and tested on independent patients in the population. The second is trained and tested on independent patients in each of $k = 3$ clusters. The latter improved performance.

Estimator	Median [IQR] error (mmHg)
Baseline	5.69 [1.47–8.30]
Clustered	3.21 [1.92–6.51]

within in cluster.

Testing error is summarized in Table 5.1. The median [IQR] error of the baseline predictor was 5.69 [1.47–8.30] mmHg, and of the clustered predictors, this was reduced to 3.21 [1.92–6.51]. The basis for this improvement can be seen in Fig. 5.3: there is minimal correlation between mean CBFV and mean ICP across the population, but there is moderate correlation within each cluster.

5.5 Extensions

5.5.1 Effect of λ

As noted in Section 5.4.1, the clustering method is robust to changes in the tradeoff constant λ. There is a reasonable explanation for this behavior. In (5.19), λ controls the tolerance for modifying the data in order to retrieve a group-sparse graph G. When λ is very large, the $\|\mathcal{O} - \mathcal{O}G\|_F^2$ penalty closely resembles the $\mathcal{O} = \mathcal{O}G$ constraint, and V will recover the 'true' clusters of models lying in the same subspace. As λ decreases, more emphasis is placed on finding a group-sparse encoding for G by allowing more deviation from the data in $\|\mathcal{O} - \mathcal{O}G\|_F^2$. So, λ plays an important role in regulating the sparsity in G and, by extension, V. This relationship is demonstrated in Fig. 5.4.

However, the actual number of *cluster assignments* is determined by the user-defined k. So, if V contains less than k subspaces, additional cuts to the graph will be made during the spectral clustering step until there are k disconnected graphs remaining. Because of

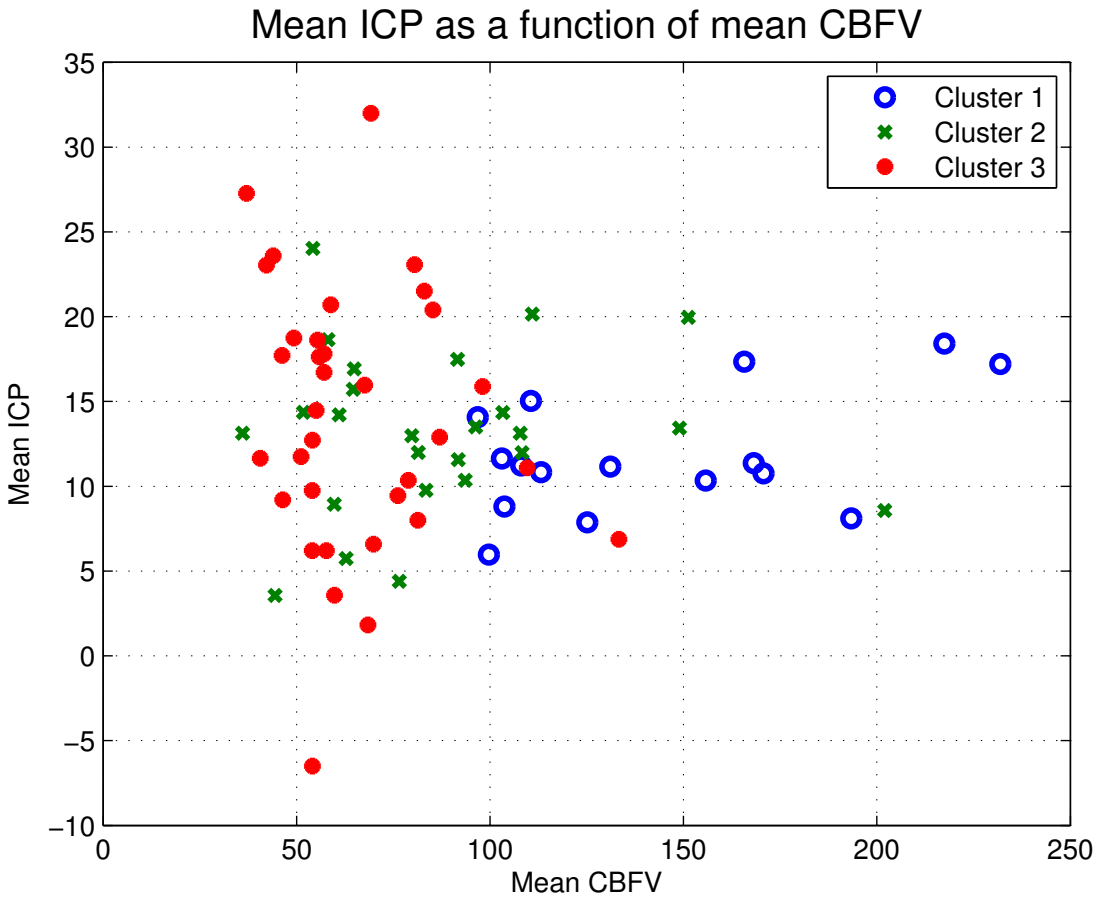

Figure 5.3: The relationship between mean CBFV and mean ICP is in general poorly correlated. However, the intra-cluster correlations reveal moderately improved correlations. This result suggests that applying system clustering as a preprocessing step to NICP estimation can improve outcomes.

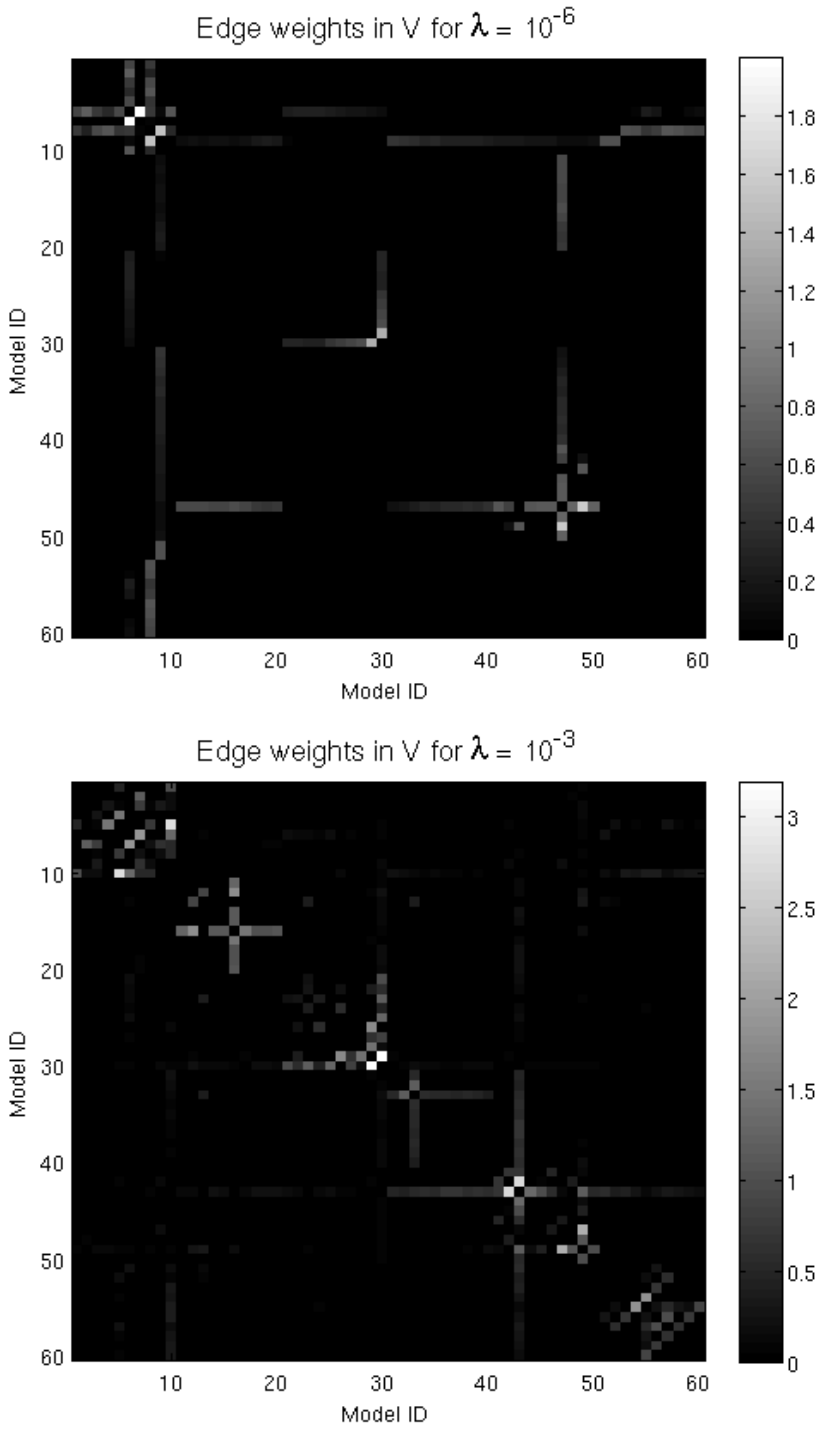

Figure 5.4: Effect of λ on the sparsity pattern of V for $L = 60$ models drawn from 6 patients.

this process, the choice of k typically dominates the choice of λ. However, if λ is extremely small, the graph may be too sparse to find a reasonable reconstruction, producing undesired results.

It is worth noting that the graph V is useful for understanding the relationship between the clustered patient models. The weights may also be used directly to describe similarities between patients without assigning Boolean cluster memberships. In these cases, the choice of λ plays a more important role in the problem.

5.5.2 Choice of group sparsity penalty

In problem (5.19), the sum-of-Frobenius-norms penalty

$$\sum_{i=1}^{L}\sum_{j=1}^{L}\|G_{ij}\|_F \tag{5.22}$$

is the Schatten (matrix) norm analogue of the ℓ_1/ℓ_2-norm that is used to promote group sparsity in vectors. Although this is a reasonable default choice, there are other interesting choices of composite norms to induce different properties into this group sparsity in G. Possible alternatives include the sum-of-spectral-norms

$$\sum_{i=1}^{L}\sum_{j=1}^{L}\|G_{ij}\|_2, \tag{5.23}$$

and the sum-of-trace-norms

$$\sum_{i=1}^{L}\sum_{j=1}^{L}\|G_{ij}\|_*. \tag{5.24}$$

It is not obvious which norm is the best choice; however, some differences can be anticipated. For example, the sum-of-trace-norms will produce *low-rank* blocks G_{ij}, which may be advantageous for certain circumstances—for example, when solving the sparse column space clustering method (5.16) for low-rank data that has been corrupted by noise. Another difference is that the sum-of-spectral-norms does not have a closed-form proximal operator, and so is a more computationally expensive choice.

For the application of ICP dynamics, comparisons between the choice of norm are made by comparing clustering accuracy. The simulations in Section 5.4.1 of $L = 100$ models are

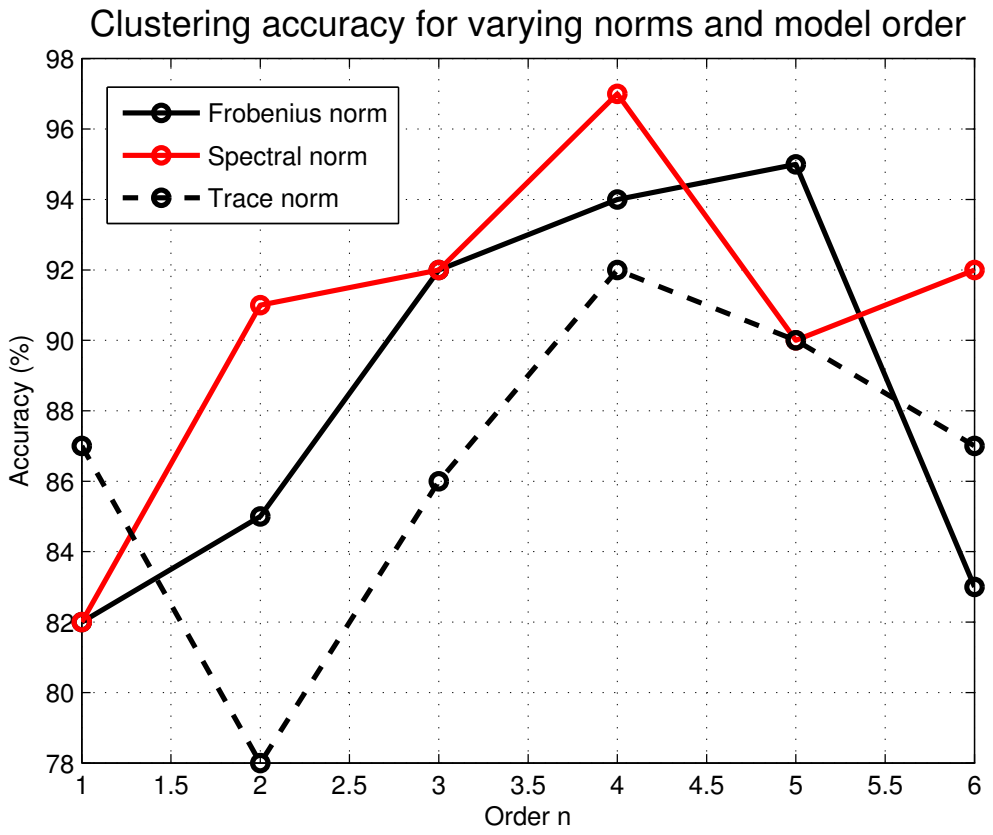

Figure 5.5: The accuracy of system clustering for three different choices of group sparsity norm penalty, and models of order $n = 1, \ldots, 6$. The $L = 100$ models were constructed from 10 different patients for this experiment. While the trace norm underperforms its alternatives, there is otherwise no strong trends in the choices. Low-order models appear to work acceptably, and require substantially less computation than higher-order models.

repeated for the above norm choices (5.22)–(5.24), and for models of order $n = 1, \ldots, 6$. Results are shown in Fig. 5.5. There is no clear distinction between the three choices, suggesting that the default option, the sum-of-Frobenius-norms, is a reasonable choice for the case of ICP dynamics.

5.5.3 Data matrix methods

Section 3.2 demonstrated that a row permutation of the block Hankel matrix

$$W^i_{r,N} = \begin{bmatrix} w^i(0) & w^i(1) & \cdots & w^i(r-1) \\ w^i(1) & w^i(2) & \cdots & w^i(r) \\ \vdots & \vdots & \ddots & \vdots \\ w^i(N-r) & w^i(N-r+1) & \cdots & w^i(N-1) \end{bmatrix} \in \mathbb{R}^{(N-r+1)d \times r} \tag{5.25}$$

could be used to estimate the column space of $\mathcal{O}_i$ of an LTI system, where $w^i(k) \in \mathbb{R}^d$ is the concatenation of observed signals from Σ_i, sampled at $k = 0, \ldots, N-1$. So, it is natural to suggest that $W^i_{r,N}$ could replace $\mathcal{O}_i$ in (5.19) to cluster these systems:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{L} \sum_{j=1}^{L} \|G_{ij}\|_F + \frac{\lambda}{2} \|W - WG\|_F^2 \\ \text{subject to} \quad & G_{ii} = 0, \quad i = 1, \ldots, L, \end{aligned} \tag{5.26}$$

where

$$W = \begin{bmatrix} W^1_{r,N} & \cdots & W^L_{r,N} \end{bmatrix}. \tag{5.27}$$

This formulation is impractical for many situations. Typically, the number of elements in $W^i_{r,N}$ is substantially larger than the number of data measurements N, which is itself large. In contrast, for a low-order system, $\mathcal{O}_i$ can be just a small number of elements. The increase in computational burden of clustering $W^i_{r,N}$ in place of $\mathcal{O}_i$ makes it an impractical alternative.

Simulations were conducted by reducing the size of the problem. Data was limited to $N = 500$ samples (12.5 s of time series data at 40 Hz), and the number of columns in each $W^i_{r,N}$ was set to $r = 3$ to significantly reduce the size of the optimization variable $G \in \mathbb{R}^{rL \times rL}$. The method was able to cluster $L = 100$ models into their correct 10 patient labels with 85% accuracy, underperforming the observability matrix approach (5.19), and being 33× slower to compute this small problem. Additional simulations that instead used the sum-of-trace-norms and sum-of-spectral-norms for group sparsity each achieved 81% accuracy. So, this variation is not in itself competitive with the original method (5.19) that was proposed.

The use of $W^i_{r,N}$ is interesting to consider for potential future use because of a special property: it is a linear function of the measurement data $w^i(k)$. So, clustering systems by the column space of this alternative matrix suggests a possible extension in which $w^i(k)$ is a variable in an optimization problem similar to (5.26). One example of how this could be useful is if the ICP signal of a new patient is treated as missing data. In such a problem, the patient would be clustered and have their ICP estimated simultaneously by constraining known elements of $w^i(k)$ to be equal to their measurement, and including a regularization term such as $\|W^i_{r,N}\|_*$ to enforce low-order dynamics.

However, there are major limitations to implementing such a method. Such an approach would result in a *nonconvex* optimization problem, even further exacerbating the computational issues of (5.26). Due to these substantial challenges and limited benefit, a tractable solution to this idea is outside of the scope of this dissertation.

5.6 Comparison to distance matrix–based methods

This section will compare the performance of the new method (5.19) to a baseline approach. Because there is no current method for general-purpose time series data, a simple clustering method based is proposed for comparison. In this approach, pairwise distances between models are used to construct distance matrices, which can be clustered using existing tools.

Two distance metrics will be considered. The first is the Martin distance, which was first given in (5.4). For two systems (Σ_i, Σ_j) forming principal angles $\theta_1, \ldots, \theta_p$, the Martin distance is

$$d_M(\Sigma_i, \Sigma_j) = \sqrt{\ln \prod_{i=1}^{p} \frac{1}{\cos^2 \theta_i}}. \tag{5.28}$$

The metric second is novel, and is inspired by (5.19):

$$d_R(\Sigma_i, \Sigma_j) = \min_G \left\{ \frac{1}{2} \|\mathcal{O}_i - \mathcal{O}_j G\|_F^2 \right\} + \min_G \left\{ \frac{1}{2} \|\mathcal{O}_j - \mathcal{O}_i G\|_F^2 \right\}. \tag{5.29}$$

This metric will be called the *subspace reconstruction distance*, and measures the residual between an attempted reconstruction of $(\mathcal{O}_i, \mathcal{O}_j)$ to the same subspace. It is a least squares

problem with the closed form solution

$$\begin{aligned} d_R(\Sigma_i, \Sigma_j) &= \frac{1}{2}\|(1 - \mathcal{P}(\mathcal{O}_j))\mathcal{O}_i\|_F^2 + \frac{1}{2}\|(1 - \mathcal{P}(\mathcal{O}_i))\mathcal{O}_j\|_F^2 \\ &= \frac{1}{2}\|\mathcal{O}_i - \mathcal{O}_j \left(\mathcal{O}_j^\top \mathcal{O}_j\right)^{-1} \mathcal{O}_j^\top \mathcal{O}_i\|_F^2 + \frac{1}{2}\|\mathcal{O}_j - \mathcal{O}_i \left(\mathcal{O}_i^\top \mathcal{O}_i\right)^{-1} \mathcal{O}_i^\top \mathcal{O}_j\|_F^2. \end{aligned} \tag{5.30}$$

Distance matrices can be used to cluster objects whose pairwise distances can be measured, even if they do not have absolute coordinates. One popular example is to use *hierarchical clustering* methods [Joh67]. In the context of clustering LTI systems, two distance matrices $D_M \in \mathbb{S}^{L\times L}$ and $D_R \in \mathbb{S}^{L\times L}$ are proposed, each defined by having the elements

$$\begin{aligned} [D_M]_{ij} &= d_M(\Sigma_i, \Sigma_j), \\ [D_R]_{ij} &= d_R(\Sigma_i, \Sigma_j). \end{aligned} \tag{5.31}$$

In a simulation, D_M and D_R were constructed for the $L = 100$ models taken from episodes from 10 patients. Clusters were constructed ($k = 10$) using MATLAB's agglomerative hierarchical clustering toolbox [Jon97]. Results of the simulation are shown in Fig. 5.6. The approach was unable to recover the underlying per-patient clusters for neither choice of distance matrix. This result provides additional evidence towards the usefulness of the new subspace-based system clustering method.

5.7 Discussion

This chapter presented a new method for clustering LTI systems by the similarity of their dynamics. It operates by comparing the column space of the systems' observability matrices, and can be computed by solving a separable convex optimization problem. Simulations validated the method and demonstrated its value.

As a standalone method, potential future work would be driven by new applications. Such cases may facilitate or necessitate new extensions to the clustering approach that do not arise in the NICP estimation problem. One example is the concept of imputing missing values during clustering, which is infeasible for the purposes of this dissertation.

In the next chapter, this new framework will be applied to the NICP problem. New

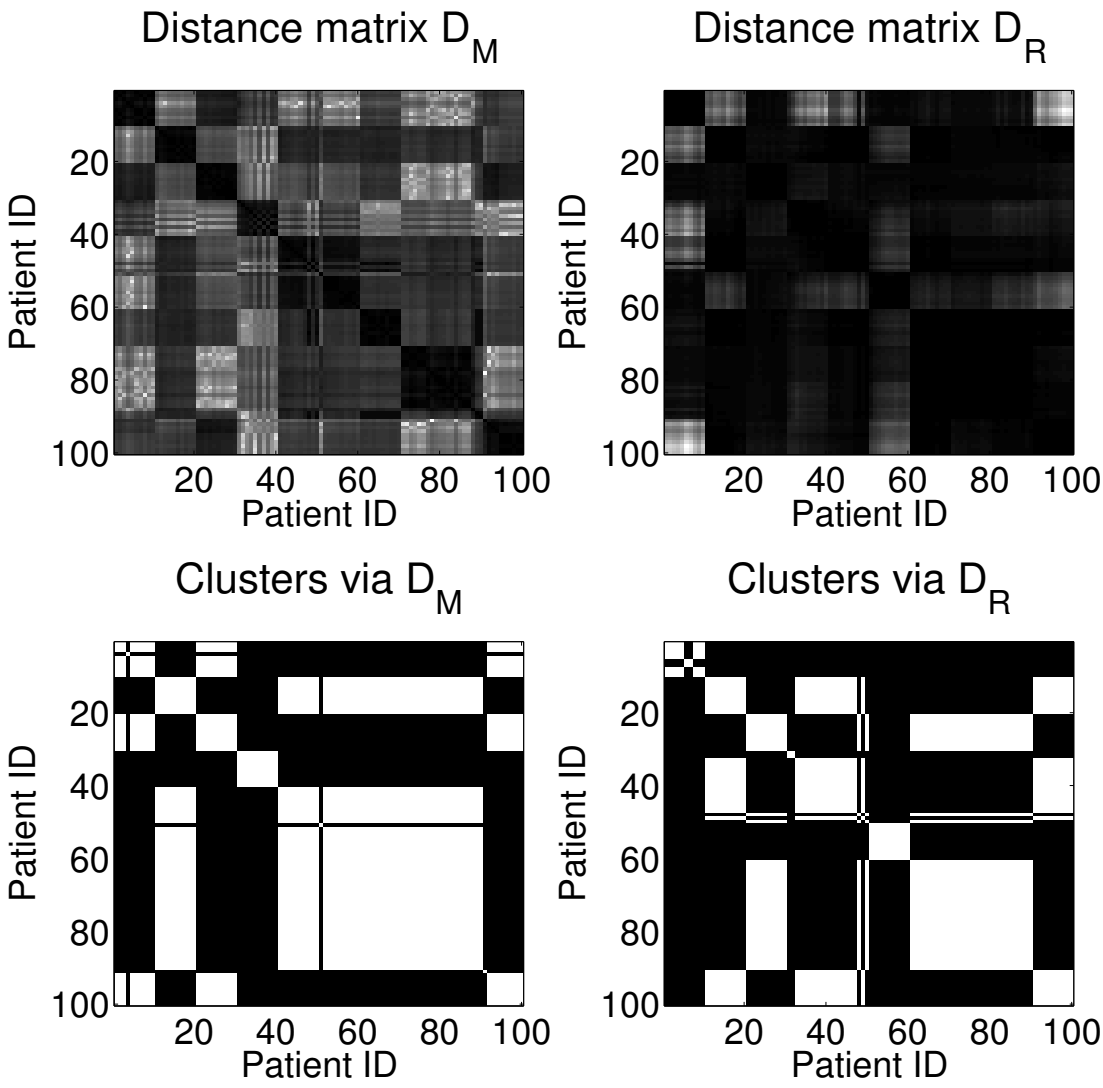

Figure 5.6: Distance matrices D_M and D_R, and the corresponding clusters that were assigned via agglomerative hierarchical clustering. Neither of the cluster assignments is able to recover the underlying per-patient clusters, severely underperforming the method of (5.19) based on sparse column space clustering.

patients will be clustered with with a dictionary using noninvasive data to find a subset of the population with similar dynamics to base NICP estimates on.

CHAPTER 6

Dictionary-based signal estimation with application to NICP estimation

6.1 Introduction

This chapter presents a new method for blind signal estimation called *dictionary-based signal estimation* (DBSE). The problem it solves was first stated in Section 1.1; it is repeated here for convenience:

> A patient's ICP is an unknown, time-varying signal. The aim of the task is to estimate this signal. All that is measured of a patient are two 'noninvasive' time-varying signals: their *cerebal blood flow velocity* (CBFV), and their *arterial blood pressure* (ABP). It is assumed that the relationship between ABP, CBFV, and ICP can be accurately modeled by a low-order (that is, low-complexity) linear dynamical system, in which ABP and CBFV are the inputs, and ICP is the output. However, this system is unknown. Because only the system inputs are measured, and the system and its output are both unknown, this is a *blind signal estimation* problem that requires additional information or assumptions to solve.
>
> Also available is a *dictionary* of past patient information. These patients had their ICP measured by invasive means, and so a complete picture is available: their input data, output (ICP) data, and system model are all known. So, the NICP problem is to combine (a) the incomplete information on a patient and (b) the complete information of other patients.

This chapter will first review other model-based NICP estimation schemes in literature. Then, the DBSE method will be presented. Finally, the method is simulated on ICP data.

6.1.1 Background on current methods

Mapping functions between models. A family of methods by Schmidt et al. were first proposed in [SKS97] and further developed in [SCS00, SCR03, SWS16]. This approach considers two single-input-single-output models. The first has the configuration ABP→CBFV (this is the same configuration as the *CBFV model* used in this dissertation), and the second model has the configuration ABP→ICP. Both models are implemented as moving average (MA) processes, that is, the output is simply a linear combination of a finite number of past inputs.

A central component of the authors' approach is that they compute a *mapping function* from training data that maps the coefficients of a patient's ABP→CBFV model to those of their ABP→ICP model. Because the ABP→CBFV model can be fit to noninvasively-measured patient data, this mapping function can be used to estimate a patient-specific ICP model. The authors reported a mean absolute error of these resulting models to be 4.1 ± 2.2 mmHg [SCS00].

Circuit analogy. Another set of methods was developed by Kashif et al. [KHV08, KVN12], which are based on the ICP modeling work of Ursino & Lodi [UL97, Urs88a, Urs88b]. These models describe the physiology of ICP dynamics using an electrical circuit analogy. Circuit parameters are fit and ICP, modeled as a voltage, is estimated from ABP and CBFV data in a single step. This approach is distinct from other NICP estimators in that it explicitly models physiological processes, rather than using a black box LTI model determined purely by the input–output data. The method has reported a mean (bias) error of 1.5 ± 5.9 mmHg.[1]

[1]Note that this quantity is distinct from mean *absolute* error.

Machine learning approaches. A final family of interesting methods has been developed by Hu et al. using a variety of machine learning approaches. These approaches include using nonlinear regression to describe the dependency of ABP and CBFV as predictors [HNG06]; kernel spectral regression and support vector machines [XKB10]; autonomously extracting important morphological features from waveforms [HXS09]; and supervised and semi-supervised techniques for binary classification of intracranial hypertension [KHP13]. Additionally, a data mining approach has been presented that creates a mapping function to provide a means to select 'similar' patient models from a dictionary [HNB06, HXW10]. These methods have resulted in mean absolute ICP errors of around 6.0 ± 5.5 mmHg, and intracranial hypertension predictions with 92% accuracy.

6.2 Dictionary-based signal estimation

6.2.1 The DBSE framework

Consider an input signal $\hat{u}_{\text{meas}}(k) \in \mathbb{R}^m$ with measurements at $k = 0, 1 \ldots, N-1$. This input is applied to an unknown low-order LTI system $\hat{\Sigma}$ to produce an unobserved (scalar-valued) output signal $\hat{y}(k) \in \mathbb{R}$. Available is a *dictionary* of L systems $\Sigma_1, \ldots, \Sigma_L$ that are similar to $\hat{\Sigma}$, each with the same number of inputs and a single output. The systems need not be of the same order. The task is to estimate the missing output $\hat{y}(k)$. An overview of this problem setup is shown in Fig. 6.1.

The dictionary-based signal estimation (DBSE) method to solve this task is posed as a convex optimization problem that minimizes the tradeoff of two terms. The first term penalizes outputs $\hat{y}(k)$ that produce high-order systems. The second term penalizes outputs that deviate from the behavior of the dictionary systems $\Sigma_1, \ldots, \Sigma_L$.

The basis for the first term is the assumption that a low-order LTI system can accurately model the relationship between $\hat{u}_{\text{meas}}(k)$ and $\hat{y}(k)$. Because $\hat{u}_{\text{meas}}(k)$ is known, the estimate for $\hat{y}(k)$ is restricted to the set of signals that could be produced by low-order dynamics. Specifically, it is expected that $\|F(\hat{u}_{\text{meas}}, \hat{y})\|_*$ is small, where again $F(\cdot, \cdot)$ is the stacked

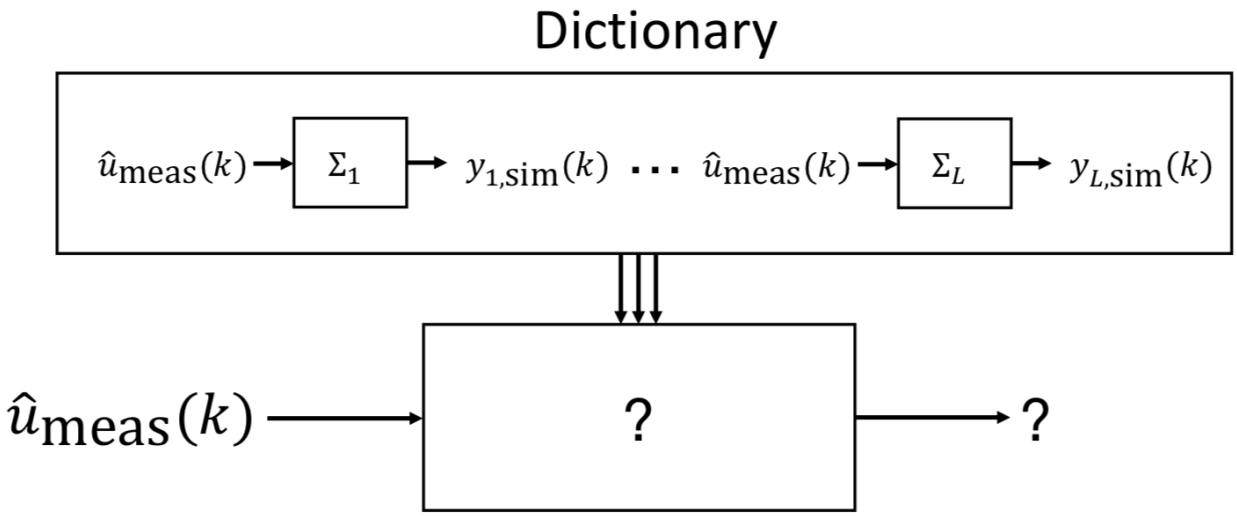

Figure 6.1: An overview of the dictionary-based signal estimation problem setup. Input signals $\hat{u}_{\text{meas}}(k) \in \mathbb{R}^m$ are measured, but the system $\hat{\Sigma}$ and the output $\hat{y}(k) \in \mathbb{R}$ are not. A dictionary of L similar systems $\Sigma_1, \ldots, \Sigma_L$ with complete information is available. Estimates of these systems' outputs can be generated by subjecting them to the known input $\hat{u}_{\text{meas}}(k)$; these simulated outputs are each denoted $y_{i,\text{sim}}(k)$.

block Hankel matrix of the form

$$F(\hat{u}, \hat{y}) = \begin{bmatrix} \hat{U}_{r,N} \\ \hat{Y}_{r,N} \end{bmatrix}, \tag{6.1}$$

that satisfies the persistency of excitation condition in (3.8).

The second term minimized in the optimization problem is based on the assumption that the system $\hat{\Sigma}$ exhibits similar dynamics to those in the dictionary, $\Sigma_1, \ldots, \Sigma_L$. So, if these dictionary systems are subjected to the same input $\hat{u}_{\text{meas}}(k)$, then their outputs should be similar to $\hat{y}(k)$. This idea can be realized in two steps. First, the input measurement $\hat{u}_{\text{meas}}(k)$ is applied to each system Σ_i to produce some simulated output $y_{i,\text{sim}}(k)$. Then, the target signal $\hat{y}(k)$ is penalized by its deviation from each $y_{i,\text{sim}}(k)$.

If each system Σ_i has a state space representation with matrices (A_i, B_i, C_i, D_i), then its dictionary simulation $y_{i,\text{sim}}(k)$ can be evaluated by

$$y_{i,\text{sim}}(k) = C_i A_i^k x(0) + \left[\sum_{j=0}^{k-1} C_i A_i^{k-1-j} B_i \hat{u}_{\text{meas}}(j) \right] + D_i \hat{u}_{\text{meas}}(k). \tag{6.2}$$

In the case of zero initial conditions, that is, $x(0) = 0$, then $y_{i,\text{sim}}(k)$ can be solved a priori from the known data as

$$y_{i,\text{sim}}(k) = \left[\sum_{j=0}^{k-1} C_i A_i^{k-1-j} B_i \hat{u}_{\text{meas}}(j) \right] + D_i \hat{u}_{\text{meas}}(k), \tag{6.3}$$

and then treated as constant in the steps that follow. (See Section 6.2.3 for a discussion on how to handle initial conditions otherwise). For compactness, the following vector notation will be used:

$$\hat{y} = \begin{bmatrix} \hat{y}(0) \\ \vdots \\ \hat{y}(N-1) \end{bmatrix} \in \mathbb{R}^N, \qquad y_{i,\text{sim}} = \begin{bmatrix} y_{i,\text{sim}}(0) \\ \vdots \\ y_{i,\text{sim}}(N-1) \end{bmatrix} \in \mathbb{R}^N. \tag{6.4}$$

There are numerous convex penalties $h(\hat{y} - y_{i,\text{sim}})$ that can be applied to the output deviations, for example the ℓ_2-norm, the ℓ_1-norm, and the Huber loss [Hub64]. Furthermore, there are multiple ways to collect the error terms for each of the L systems. One simple

choice is the weighted-sum-of-penalties

$$\sum_{i=1}^{L} \gamma_i^2 h(\hat{y} - y_{i,\text{sim}}),$$

with weights $\gamma_i > 0$ and $\sum_{i=1}^{L} \gamma_i = 1$. Another choice is the maximum penalty:

$$\max_{i=1,\ldots,L} \left(h(\hat{y} - y_{i,\text{sim}}) \right).$$

Combining these two terms to form an optimization problem, using the weighted-sums-of-penalties example, the DBSE method is

$$\text{minimize} \qquad \|F(\hat{u}_{\text{meas}}, \hat{y})\|_* + \frac{\zeta}{2} \sum_{i=1}^{L} \gamma_i^2 h(\hat{y} - y_{i,\text{sim}}), \tag{6.5}$$

where $\zeta > 0$ is a tradeoff constant. One interesting example arises when $h(x) = \|x\|_2^2$, that is,

$$\text{minimize} \qquad \|F(\hat{u}_{\text{meas}}, \hat{y})\|_* + \frac{\zeta}{2} \sum_{i=1}^{L} \gamma_i^2 \|\hat{y} - y_{i,\text{sim}}\|_2^2. \tag{6.6}$$

This problem is equivalent to the more compact expression

$$\text{minimize} \qquad \|F(\hat{u}_{\text{meas}}, \hat{y})\|_* + \frac{\zeta}{2} \|\hat{y} - y_{\text{sim}}\|_2^2, \tag{6.7}$$

in which y_{sim} is the weighted average of the dictionary simulations,

$$y_{\text{sim}} = \sum_{i=1}^{L} \gamma_i \; y_{i,\text{sim}}, \tag{6.8}$$

with $\sum_{i=1}^{L} \gamma_i = 1$. This simplified formulation will be used in the sections that follow.

When ζ is large, the problem (6.7) tends towards choosing $\hat{y} = y_{\text{sim}}$, the weighted average of the dictionary simulations. As ζ is reduced, the problem produces solutions $\hat{y}$ that tend towards a lower-order relationship with the input data. In this sense, $\|F(\hat{u}_{\text{meas}}, \hat{y}))\|_*$ acts as an approximate low-order regularization term to $\hat{\Sigma}$. A summary of this basic version of the DBSE method is given in Algorithm 2.

Algorithm 2 Dictionary-based signal estimation, basic case

1: Compute y_{sim} by subjecting the dictionary systems to $\hat{u}_{\text{meas}}(k)$ via (6.3) and the determining a weighted-average using (6.8).

2: Solve the optimization problem (6.7) to recover an estimate for $\hat{y}$.

6.2.2 Choosing simulation weights

DBSE methods such as (6.7) require the user to choose weights γ_i that regulate the influence of each system Σ_i on the simulated output y_{sim}. This choice of weights is important, and a simple uniform weight strategy (i.e. $\gamma_i = 1/L$ for $i = 1, \ldots, L$) is unlikely to be the best choice for practical applications. A better approach is to determine, a priori, which dictionary systems have most similar dynamical properties to $\hat{\Sigma}$, and weight them accordingly. This approach typically requires additional assumptions on the available data.

Target systems and related systems. Consider a low-order LTI system Σ_i, which is called the *target system*. The output of the target system is $y(k) \in \mathbb{R}$. If $y(k)$ is omitted from the dataset, and the remaining signals can still be accurately modeled by a low-order LTI system, then this new system is labeled Γ_i and is called the *related system*.[2]

This scenario arises in NICP estimation. Here, the CBFV model is the *related system* Γ, with the input–output structure

$$u(k) = \text{ABP}(k), \qquad y(k) = \text{CBFV}(k), \tag{6.9}$$

and the ICP model is the *target system* Σ, with structure

$$u(k) = \begin{bmatrix} \text{ABP}(k) \\ \text{CBFV}(k) \end{bmatrix}, \qquad y(k) = \text{ICP}(k). \tag{6.10}$$

The accuracy of these relationships as low-order LTI models was verified in Section 4.2.

This property is important for choosing the simulation weights γ. Although the target system $\hat{\Sigma}$ is unknown, and so cannot be compared to $\Sigma_1, \ldots, \Sigma_L$, the related system $\hat{\Gamma}$ can be

[2]These names are inspired by the *target time series* and *related time series* used in [HNB06, KSB12].

fit to the available data, and compared to $\Gamma_1, \ldots, \Gamma_L$. One idea for comparing these related systems is to use the system clustering methods introduced in Chapter 5.

System clustering. As suggested, system clustering can be applied to the related systems $(\hat{\Gamma}, \Gamma_1, \ldots, \Gamma_L)$. To do this, these systems are first fit to the available time series data using a routine system identification tool (e.g. N4SID [VD94]). Next, the *system clustering* method described in Algorithm 1 is applied to the $L+1$ systems $(\hat{\Gamma}, \Gamma_1, \ldots, \Gamma_L)$. This step requires the user to choose the number of clusters k. This process produces a subset of the dictionary systems Γ_i that are assigned to the same cluster as $\hat{\Gamma}$, and are expected to exhibit the most-similar dynamical behavior.

A simple way to use this dictionary subset for weight selection is to uniformly include only those dictionary systems that are within the same cluster as $\hat{\Gamma}$. For example, if $\hat{L}$ elements reside in the same cluster as $\hat{\Gamma}$, then

$$\gamma_i = \begin{cases} 1/\hat{L}, & \Gamma_i \text{ clustered with } \hat{\Gamma} \\ 0, & \text{else} \end{cases} . \tag{6.11}$$

Another alternative to this scheme is to weight each γ_i by the edge weight between $(\hat{\Gamma}, \Gamma_i)$, which given in the matrix V in (5.17), and then normalizing these weights such that $\sum_{i=1}^{L} \gamma_i = 1$.

Other approaches. Numerous other schemes for weighting the dictionary simulations are possible, particularly by comparing each pairing of the related systems $(\hat{\Gamma}, \Gamma_i)$. Ideas include weighting by the Martin distance between the systems [DD02], or developing an error mapping function between systems to measure their expected simulation error, as in [KSB12].

6.2.3 Extensions

Initial conditions. The method (6.7) assumes that each simulation $y_{i,\text{sim}}(k)$ has zero initial conditions, that is, $x(0) = 0$. This is unrealistic for most cases, and can result in a

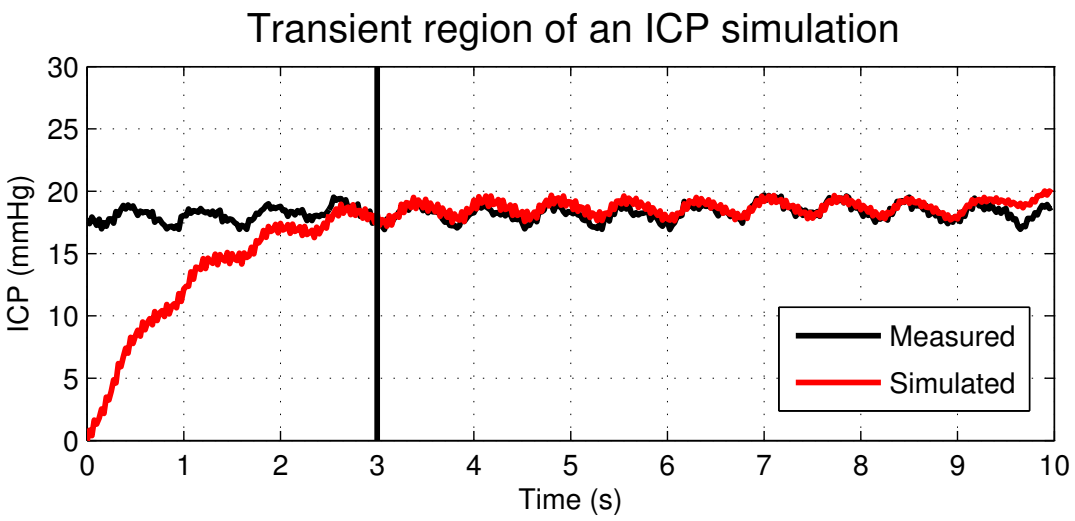

Figure 6.2: Simulation of an ICP signal, assuming zero initial conditions. The transient region $t \in (0, 3)$ s is identified and excluded when solving (6.7).

transient response period before the system settles to a *steady state response* [FPE94]. Perhaps the simplest way to overcome this issue is to simulate $y_{i,\text{sim}}(k)$ without initial conditions, identify the transient response region of this simulation, and then exclude this section of data when solving the optimization problem (6.7). For an example, see Fig. 6.2. This is a practical solution because it preserves the ability to treat y_{sim} as a constant in the optimization problem. Simulations in this chapter will take this approach.

A more sophisticated approach is to fit, for each simulation $y_{i,\text{meas}}(k)$, an optimal choice of $x(0)$. For example, consider the problem

$$\text{minimize} \qquad \|F(\hat{u}_{\text{meas}}, \hat{y})\|_* + \frac{\zeta}{2} \sum_{i=1}^{L} \gamma_i \|\hat{y} - y'_{i,\text{sim}}\|_2^2. \tag{6.12}$$

where $y'_{i,\text{sim}}(k)$ is the variable

$$\begin{aligned} y'_{i,\text{sim}}(k) &= C_i A_i^k x(0) + \left[\sum_{j=0}^{k-1} C_i A_i^{k-1-j} B_i \hat{u}_{\text{meas}}(j) \right] + D_i \hat{u}_{\text{meas}}(k) \\ &= C_i A_i^k x(0) + y_{i,\text{sim}}(k). \end{aligned} \tag{6.13}$$

That is, $y'_{i,\text{sim}}(k)$ is the sum of a transient term $C_i A_i^k x(0)$, in which $x(0)$ is a variable, and the original term $y_{i,\text{meas}}(k)$, which can be computed a priori via (6.3) and treated as a constant. So, solving (6.12) then additionally fits an optimal $x(0)$ for each of the L dictionary

simulations that further minimizes the difference between $\hat{y}(k)$ and $y_{i,\text{sim}}(k)$, in the ℓ_2-norm sense.

Corrupted inputs. If the input data $u_{\text{meas}}(k)$ is corrupted, it may not be desirable to solve (6.7) with exact data. This may be the case in NICP estimation, because the input CBFV signals can be corrupted, an issue that was discussed in Chapter 4. A simple way to overcome this issue is to remove the effects of these corruptions a priori. For example, simulations in Section 4.4 demonstrated that the methods in this dissertation can accurately remove these artifacts and impute missing sections using only the routinely-measured ABP and CBFV data.

However, these approaches can also be integrated into the DBSE problem in (6.7). In one scenario, if the input data is exact but some measurements are missing, these missing elements can simply be omitted from the constraints. If elements that were measured in each scalar signal $\hat{u}_{i,\text{meas}}(k)$ are indexed by the index set $\mathcal{K}_i \subseteq \{0, \ldots, N-1\}$, for $i = 1, \ldots, m$, then this becomes

$$\begin{aligned} \text{minimize} \quad & \|F(\hat{u}, \hat{y})\|_* + \frac{\zeta}{2}\|\hat{y} - y_{\text{sim}}\|_2^2 \\ \text{subject to} \quad & \hat{u}_i(k) = \hat{u}_{i,\text{meas}}(k), \quad i = 1, \ldots, m, \quad k \in \mathcal{K}_i. \end{aligned} \tag{6.14}$$

The missing elements in $\hat{u}_{\text{meas}}(k)$ will be imputed with values $\hat{u}(k)$ that keep the system order low.

If the data contains other corruptions, the constraints can again be replaced with a penalty function, for example

$$\text{minimize} \quad \|F(u, y)\|_* + \frac{\zeta}{2}\|\hat{y} - y_{\text{sim}}\|_2^2 + \frac{\rho}{2}\sum_{i=1}^{m}\sum_{k\in\mathcal{K}_i}\|\hat{u}_i(k) - \hat{u}_{i,\text{meas}}(k)\|_1, \tag{6.15}$$

or any other useful choice of loss function on $\hat{u}(k) - \hat{u}_{\text{meas}}(k)$. In this scenario, $y_{\text{sim}}(k)$ is a linear function of $\hat{u}(k)$ and so it is also a variable. Caution should be taken with this approach: unless $\rho \gg \zeta$, the *simulated* output data from the other dictionary systems may provide more influence to the solution than the *measured* inputs of the current system, which could lead to undesired outcomes.

6.2.4 Algorithm

The central step to the DBSE method is to solve the convex optimization problem (6.7). This problem does not have a closed-form solution, but can be solved efficiently using proximal algorithms. The steps to apply such algorithms to the DBSE method are covered in Section 7.6.

6.3 Simulations

In this section, the DBSE method is used to design a simple NICP estimator. In this scenario, the target system $\hat{\Sigma}$ represents a patient whose ABP and CBFV have been measured, but their ICP, the output, is unknown. The $L = 75$ dictionary systems $\Sigma_1, \ldots, \Sigma_L$ represent past patients whose ICP was invasively measured via IVC. Systems have the structure of the ICP model, that is,

$$u(k) = \begin{bmatrix} \text{ABP}(k) \\ \text{CBFV}(k) \end{bmatrix}, \quad y(k) = \text{ICP}(k). \tag{6.16}$$

The inputs to the current patient model $\hat{\Sigma}$, $\hat{u}_{\text{meas}}(k)$, are known because they were measured by noninvasive means. The related system $\hat{\Gamma}$ has the structure of the CBFV model: $u(k) =$ ABP and $y(k) =$ CBFV, a model that can be constructed using noninvasive signals only.

The DBSE method (6.7) is applied to this data to estimate patients' ICP. Three different weighting schemes are used to choose γ. The first is a uniform scheme, $\gamma_i = 1/L$ for $i = 1 \ldots, L$. For the second scheme, the related system models $(\hat{\Gamma}, \Gamma_1, \ldots, \Gamma_L)$ are first clustered. The dictionary systems that are assigned to the same cluster as $\hat{\Gamma}$ are then uniformly weighted. (See Section 6.2.2). The third scheme similarly first clusters $(\hat{\Gamma}, \Gamma_1, \ldots, \Gamma_L)$, and instead selects each γ_i proportional to the edge weights in V between $(\hat{\Gamma}, \Gamma_i)$. The parameters k, λ, and ζ are chosen using a training dictionary independent of the test dictionary results. An example NICP estimate is provided in Fig. 6.3.

Simulation results are shown in Table 6.1. The uniform weighting scheme had a median [IQR] absolute error of 5.05 [2.86–10.57] mmHg. Both clustering-based weighting schemes

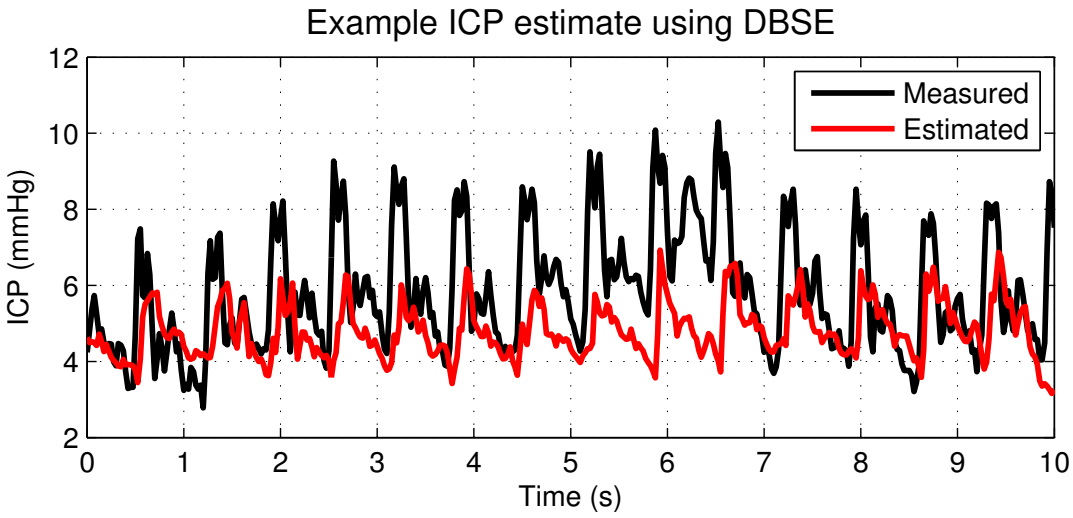

Figure 6.3: Example of an NICP estimate using the DBSE method.

Table 6.1: Median [IQR] absolute error of ICP estimates of the DBSE method in the testing set, for different weighting schemes.

Weighting	Error (mmHg)
Uniform	5.05 [2.86–10.57]
Clustered	4.71 [2.28–8.17]
V-weighted	4.54 [2.60–9.24]

improved NICP estimation error, with 4.71 [2.28–8.17] mmHg for uniform in-cluster weights, and 4.54 [2.60–9.24] mmHg when γ was set proportional to the edge weights in the graph matrix V. These results are comparable to literature, but do not offer any decisive improvement.

6.4 Discussion

This chapter has presented the *dictionary-based signal estimation* (DBSE) method, which can be used to estimate the output of an unknown LTI system if a dictionary of similar systems is available. The method operates by considering simulated outputs that would have been produced by each of the dictionary systems, with an additional regularization

term that promotes a low-order system.

Simulations demonstrated that the DBSE produced NICP estimators with comparable performance to what is in the literature. While no distinctive improvement was made over the state of the art, this approach has value in its flexible formulation. The experiments also showed that the choice of dictionary simulation weights is important, and results were improved by first clustering patient models by their noninvasive signal dynamics. So, the promise of this method is most likely as a modular component of a larger, more complex NICP framework.

CHAPTER 7

Algorithms

7.1 Introduction

Throughout this dissertation, data methods have been presented that require *convex optimization problems* to be solved in order to compute a solution. This chapter provides the algorithmic details necessary to solve these problems. In particular, the problems will be solved using *proximal algorithms*, a family of convex optimization algorithms. First, the necessary background material will be discussed. Then, three specific algorithms are described: the *alternating direction method of multipliers* (ADMM), the *primal–dual Douglas–Rachford* (PDDR) algorithm, and the *Chambolle–Pock* algorithm. Finally, details are provided to compute the solutions to the data methods presented in Chapters 4–6.

Solving convex optimization problems is in general a mature field [BV04], and many problems can be easily solved using off-the-shelf software with minimal experience. Examples of such packages include CVX [GB] and Mosek [Mos10]. Convex optimization problems are frequently solved with *second-order methods*, which use both the first and second derivative of the objective function at each iteration. Second-order methods such as Newton's method have been very successful, solving problems with high precision in just a few iterations [BV04]. However, the cost of calculating the second derivative matrix or *Hessian* becomes increasingly prohibitive for problems with a large number of variables. Therefore research interest has moved to *first-order methods*, which rely on only first derivatives, to instead solve the problem in many inexpensive steps [RB16, PB13].

Off-the-shelf optimization tools are unsuitable for the methods in this dissertation. This is because the problems presented are *large-scale*, with a significant number of variables,

and so need to be treated deliberately with first-order methods that exploit their structure to ensure tractable computation. In particular, the optimization problems in Chapters 4–6 have a consistently special form:

$$\text{minimize} \qquad f(x) + g(Ax), \tag{7.1}$$

where A is a linear transformation expressed as a matrix, and f and g are convex functions that have simple *proximal operators*. Such problems can be solved efficiently using *Douglas–Rachford splitting* algorithms.

7.2 Review of proximal operators

Proximal algorithms are a class of first-order convex optimization algorithms that use proximal operators to compute their descent step. Proximal algorithms are particularly useful for minimizing objective functions with nondifferentiable terms. This section will define proximal operators and introduce some of their useful properties.

7.2.1 Definition and interpretation

The *proximal operator* of a closed proper convex function f,

$$\text{prox}_{tf} : \mathbb{R}^n \to \mathbb{R}^n, \tag{7.2}$$

is defined

$$\text{prox}_{tf}(v) = \underset{x}{\text{argmin}} \left\{ f(x) + \frac{1}{2t}\|x - v\|_2^2 \right\}, \tag{7.3}$$

where $t > 0$ is a constant [Mor65, Roc76]. Proximal operators can be interpreted in different ways. Simply from the definition it can be seen that the proximal operator of f is some tradeoff between the minimum of f and argument v, regulated by the parameter t. One useful perspective of proximal operators is as a *generalized projection*. The proximal operator of the indicator function of a convex set $\mathcal{C}$,

$$f(x) = \mathcal{I}_{\mathcal{C}}(x) = \begin{cases} 0, & x \in \mathcal{C} \\ +\infty, & x \notin \mathcal{C} \end{cases}, \tag{7.4}$$

is the Euclidean projection onto the set:

$$\text{prox}_{tf}(v) = \underset{x \in \mathcal{C}}{\text{argmin}} \, \|x - v\|_2^2. \tag{7.5}$$

Parikh and Boyd [PB13] also note that under some additional assumptions, the proximal operator has the approximation

$$\text{prox}_{tf}(v) \approx v - t\nabla f(v), \tag{7.6}$$

for small t, which suggests that it also acts like type of gradient descent step.

7.2.2 Useful properties of proximal operators

There is a well-established body of properties that can be used to manipulate proximal operators. This section reviews some important ones, which can be used to quickly evaluate the proximal operators used in this dissertation. More detailed explanations can be found in [PB13].

Pre- and post-composition The pre-composition property states that if $f(x) = \phi(ax + b)$, then

$$\text{prox}_{tf}(v) = \frac{1}{a}\left(\text{prox}_{a^2 t\phi}(av + b) - b\right). \tag{7.7}$$

Similarly, the post-composition property states that if $f(x) = a\phi(x) + b$, then

$$\text{prox}_{tf}(v) = \text{prox}_{at\phi}(v). \tag{7.8}$$

Separability. Proximal operators of separable functions are also separable. That is, if

$$f(x) = \sum_{i=1}^{n} f_i(x_i), \tag{7.9}$$

then each proximal operator can be calculated independently as

$$\text{prox}_{tf}(x) = \begin{bmatrix} \text{prox}_{tf_1}(x_1) \\ \text{prox}_{tf_2}(x_2) \\ \vdots \\ \text{prox}_{tf_n}(x_n) \end{bmatrix}. \tag{7.10}$$

The separability property can allow complex choices of f to retain simple proximal operators.

Moreau decomposition. An important property that relates f to its conjugate function

$$f^*(y) = \sup_x \left(y^\top x - f(x) \right), \tag{7.11}$$

is the Moreau decomposition [PB13]:

$$x = \text{prox}_{tf}(x) + t\text{prox}_{t^{-1}f^*}(t^{-1}x). \tag{7.12}$$

A useful application of this result is that if the proximal operator of f is known, then it is easy to compute the proximal operator of f^*:

$$\text{prox}_{tf^*}(x) = x - t\text{prox}_{t^{-1}f}(t^{-1}x). \tag{7.13}$$

These conjugate functions appear frequently in proximal methods.

7.2.3 Examples of proximal operators

Because proximal operators involve a minimization step, it is important to identify functions that admit closed-form solutions, so that they can be iterated over efficiently. Some examples that will be used in this chapter are presented here.

Indicator function. As previously mentioned, an indicator function onto the set $\mathcal{C}$ has the operator

$$f(x) = \mathcal{I}_\mathcal{C}(x) = \begin{cases} 0, & x \in \mathcal{C} \\ +\infty, & x \notin \mathcal{C} \end{cases} \quad \Longrightarrow \quad \text{prox}_{tf}(v) = \Pi_\mathcal{C}(v) = \underset{x \in \mathcal{C}}{\text{argmin}} \, \|x - v\|_2^2. \tag{7.14}$$

Quadratic function. For a quadratic function with $A \in \mathbb{S}_+^n$,

$$f(x) = \frac{1}{2} x^\top A x + b^\top x + c \quad \Longrightarrow \quad \text{prox}_{tf}(v) = (I + \lambda A)^{-1}(v - \lambda b). \tag{7.15}$$

ℓ_1-norm. The proximal operator of an ℓ_1-norm is

$$f(x) = \|x\|_1 = \sum_{i=1}^n |x_i| \quad \Longrightarrow \quad \left(\text{prox}_{tf}(v)\right)_i = \begin{cases} v_i - t, & v_i \geq t \\ 0, & |v_i| \leq t \\ v_i + t, & v_i \leq -t \end{cases}, \tag{7.16}$$

for $i = 1, \ldots, n$. This function is called *soft thresholding* [Don95], and is shown in Fig. 7.1.

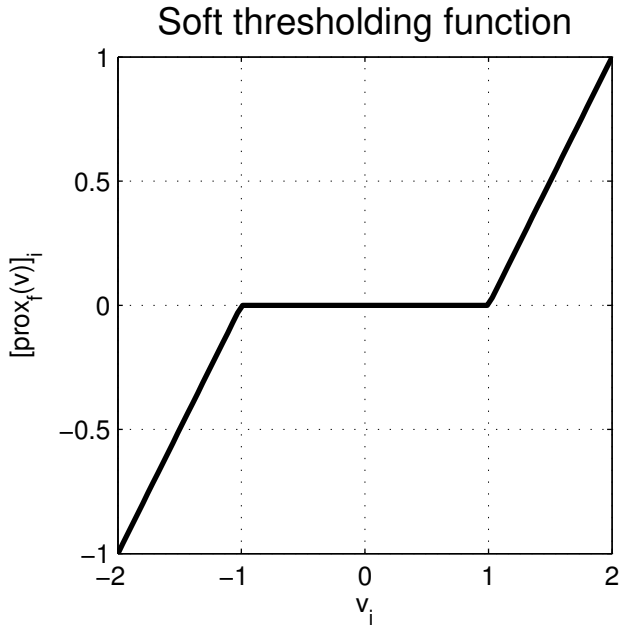

Figure 7.1: Shape of the soft thresholding function, which is the element-wise proximal operator for the ℓ_1-norm.

Trace norm. The final proximal operator of interest is for the trace norm, with $X \in \mathbb{R}^{m \times n}$:

$$f(X) = \|X\|_* \quad \Longrightarrow \quad \text{prox}_{tf}(V) = \sum_{i=1}^{\min(m,n)} (\sigma_i - t)_+ u_i v_i^\top, \quad \text{for } V = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^\top, \tag{7.17}$$

where

$$(\sigma_i - t)_+ = \begin{cases} 0, & \sigma_i - t \leq 0 \\ \sigma_i - t, & \sigma_i - t \geq 0 \end{cases}. \tag{7.18}$$

To interpret this operator, recall that the trace norm is the sum of the singular values of a matrix

$$\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i, \tag{7.19}$$

and that these values are nonnegative $\sigma_i \geq 0$. Hence, $\|X\|_* = \|\sigma\|_1$. As expected, (7.17) can then be understood as the soft thresholding function, applied to the singular values σ. Note that this proximal operator requires the singular value decomposition of its argument, which can be computationally expensive for large matrices V.

7.3 Review of Douglas–Rachford splitting algorithms

Douglas–Rachford splitting algorithms [DR56, LM79, EB92] are a class of proximal algorithms that can efficiently solve the problem

$$\text{minimize} \qquad f(x) + g(Ax), \tag{7.20}$$

when f and g are convex functions with simple proximal operators and $A \in \mathbb{R}^{m \times n}$ is some linear transformation expressed as a matrix. These algorithms have found broad application in recent years, with applications to image processing [OV14, CP11], signal processing [WBA12], blind source separation [BF10], computer resource allocation [Bir14], matrix decomposition [SAJ14], dictionary learning [CB17, RMN18], matrix and tensor completion [RP13], and distributed optimization [ZLS12, ZH18].

This section overviews a selection of the splitting algorithms that are used in this dissertation. An overview of how these algorithms can be derived in general can be found in [RB16, OV14].

7.3.1 Alternating direction method of multipliers

The most widely-used algorithm in this class is the *alternating direction method of multipliers* (ADMM) [GM76]. ADMM is sometimes analyzed in the context of augmented Lagrangian methods, although it also falls into the Douglas–Rachford framework [EB92]. In the classical approach, (7.20) is rewritten as the equivalent constrained problem

$$\begin{aligned} \text{minimize} \qquad & f(x) + g(y) \\ \text{subject to} \qquad & y = Ax, \end{aligned} \tag{7.21}$$

which has the *augmented Lagrangian* function

$$L_\rho(x, y, v) = f(x) + g(y) + v^\top (y - Ax) + \frac{\rho}{2} \|y - Ax\|_2^2, \tag{7.22}$$

with dual variable v and constant $\rho > 0$. The **ADMM step** is

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_{\hat{x}} L_\rho(\hat{x}, y^k, v^k) \\ y^{k+1} &= \operatorname*{argmin}_{\hat{y}} L_\rho(x^{k+1}, \hat{y}, v^k) \\ v^{k+1} &= v^k + \rho(y^{k+1} - Ax^{k+1}). \end{aligned} \tag{7.23}$$

This algorithm converges linearly; details on convergence and parameter selection can be found in [NLR15, SLY14, WYZ15, GTS15].

The classical formulation of the algorithm in (7.23) is only efficient if the minimization steps are inexpensive, for example if they have closed-form solutions. When this is not true, one popular workaround is to use a dummy variable. This is achieved by rewriting (7.20) as

$$\begin{aligned} \text{minimize} \quad & f(x) + g(y) \\ \text{subject to} \quad & y = Az \\ & x = z, \end{aligned} \tag{7.24}$$

for which the augmented Lagrangian, with dual variables v and w, is

$$L_\rho(x, y, z, v, w) = f(x) + g(y) + v^\top(y - Az) + w^\top(x - z) + \frac{\rho}{2}\|y - Az\|_2^2 + \frac{\rho}{2}\|x - z\|_2^2. \tag{7.25}$$

For this alternative problem configuration, the ADMM algorithm is equivalent to

$$\begin{aligned} x^{k+1} &= \operatorname{prox}_{\rho^{-1}f}\left(z^k - \rho^{-1}w^k\right) \\ y^{k+1} &= \operatorname{prox}_{\rho^{-1}g}\left(Az^k + \rho^{-1}v^k\right) \\ z^{k+1} &= \left(\rho A^\top A + \rho I\right)^{-1}\left(\rho\left(A^\top y^{k+1} + x^{k+1}\right) + A^\top v^k + w^k\right) \\ v^{k+1} &= v^k + \rho\left(Az^{k+1} - y^{k+1}\right) \\ w^{k+1} &= w^k + \rho\left(x^{k+1} - z^{k+1}\right). \end{aligned} \tag{7.26}$$

This variation is useful when the proximal operators of f and g can be easily computed. Because $\rho(A^\top A + I)$ is positive definite and constant at each iteration, z^{k+1} can be computed efficiently by computing a single Cholesky factorization, and then using factor–solve method at each iteration [Van18, BV18].

7.3.2 Primal–dual Douglas–Rachford algorithm

A method that uses proximal operators directly is the primal–dual Douglas–Rachford (PDDR) algorithm [OV14]. The **PDDR step** is:

$$\begin{aligned} x^k &= \operatorname{prox}_{tf}(\theta^{k-1}) \\ z^k &= \operatorname{prox}_{tg^*}(\phi^{k-1}) \\ \begin{bmatrix} u^k \\ v^k \end{bmatrix} &= \begin{bmatrix} I & tA^\top \\ -tA & I \end{bmatrix}^{-1} \begin{bmatrix} 2x^k - \theta^{k-1} \\ 2z^k - \phi^{k-1} \end{bmatrix} \\ \theta^k &= \theta^{k-1} + \rho(u^k - x^k) \\ \phi^k &= \phi^{k-1} + \rho(v^k - z^k). \end{aligned} \tag{7.27}$$

This method uses the proximal operators of f and g^* directly, and so is efficient when they admit closed-form solutions and A is not prohibitively large. The matrix

$$\begin{bmatrix} I & tA^\top \\ -tA & I \end{bmatrix} \tag{7.28}$$

is constant at each iteration, and so can be factorized exactly once. This significantly reduces the computational burden of the method.

7.3.3 Chambolle–Pock algorithm

The Chambolle–Pock algorithm [CP11] is another variation that also directly employs the proximal operators of f and g^*, but does not require a linear system to be solved at each iteration. The **Chambolle–Pock step** is

$$\begin{aligned} x^k &= \operatorname{prox}_{tf}\left(x^{k-1} - tA^\top z^{k-1}\right) \\ z^k &= \operatorname{prox}_{sg^*}\left(z^{k-1} + sA(2x^k - x^{k-1})\right). \end{aligned} \tag{7.29}$$

The primal and dual step sizes, t and s respectively, can differ, so long as they satisfy $\|A\|\sqrt{st} \leq 1$ to guarantee convergence.

7.4 Artifact rejection and missing data imputation

This section demonstrates how to compute solutions to the methods of Chapter 4 for artifact rejection and missing data imputation. Many of these methods can be solved directly using the PDDR algorithm; the necessary details are provided. One exception is the method for removing *structured* artifacts, which contains a penalty without a closed-form proximal operator. This problem is solved using ADMM.

7.4.1 Basic method

The *basic method* for artifact rejection and missing data imputation, which was presented in Section 4.3.1, is

$$\text{minimize} \qquad \|W_{r,N}\|_* + \sum_{i=1}^{d} \sum_{k \in \mathcal{K}} \lambda_i |w_i(k) - w_{i,\text{meas}}(k)|. \tag{7.30}$$

To solve this problem, first the optimization variable is defined as

$$x = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix} \in \mathbb{R}^{Nd}, \tag{7.31}$$

with x_{meas} similarly defined. The problem (7.30) can be rewritten with this choice of variables as

$$\text{minimize} \qquad \|\text{mat}(Ax)\|_* + \sum_{i \text{ measured}} \hat{\lambda}_i |x_i - x_{i,\text{meas}}|. \tag{7.32}$$

In this representation, 'i measured' refers to the elements for which there exists a measurement $x_{i,\text{meas}}$, and $\hat{\lambda}$ is an *element-wise* weight vector, that is,

$$\hat{\lambda} = \begin{bmatrix} \lambda_1 \mathbb{1}_N \\ \vdots \\ \lambda_d \mathbb{1}_N \end{bmatrix} \in \mathbb{R}^{Nd}. \tag{7.33}$$

The operation $\text{mat}(Ax)$ is the linear mapping of the elements in x to the block Hankel matrix used in (7.30), expressed as a matrix multiplication followed by a reshaping operator.

Problem (7.32) is of the form

$$\text{minimize} \qquad f(x) + g(Ax), \tag{7.34}$$

where

$$f(x) = \sum_{i \text{ measured}} \hat{\lambda}_i |x_i - x_{i,\text{meas}}|, \tag{7.35}$$

$$g(y) = \|\text{mat}(y)\|_*. \tag{7.36}$$

The proximal operator for f is defined, for each element $i = 1, \ldots, Nd$, by

$$\left[\text{prox}_{tf}(v)\right]_i = \begin{cases} v_i - t\hat{\lambda}_i, & v_i > t\hat{\lambda}_i + x_{i,\text{meas}} \\ x_{i,\text{meas}}, & -t\hat{\lambda}_i + x_{i,\text{meas}} \leq v_i \leq t\hat{\lambda}_i + x_{i,\text{meas}} \\ v_i + t\hat{\lambda}_i, & v_i < -t\hat{\lambda}_i + x_{i,\text{meas}} \\ v_i, & x_{i,\text{meas}} \text{ missing} \end{cases}. \tag{7.37}$$

The proximal operator for g^* is

$$\text{prox}_{tg^*}(z) = z - t \cdot \text{vec}\left[\sum_{i=1}^{r} \left(\sigma_i - t^{-1}\right)_+ u_i v_i^\top\right], \tag{7.38}$$

given the SVD

$$\text{mat}(t^{-1}z) = \sum_{i=1}^{r} \sigma_i u_i v_i^\top, \tag{7.39}$$

where r is the number of nonzero singular values. Given these proximal operators, problem (7.30) can be then immediately solved using the PDDR steps given in Section 7.3.2.

The above approach is easily extended to other problems. For example, in the ℓ_2-norm variation

$$\text{minimize} \qquad \|W_{r,N}\|_* + \sum_{i=1}^{d} \sum_{k \in \mathcal{K}} \frac{\lambda_i}{2} \left(w_i(k) - w_{i,\text{meas}}(k)\right)^2, \tag{7.40}$$

where the corresponding penalty function

$$f(x) = \sum_{i \text{ measured}} \frac{\hat{\lambda}_i}{2} \left(x_i - x_{i,\text{meas}}\right)^2, \tag{7.41}$$

has the proximal operator

$$\left[\text{prox}_{tf}(v)\right]_i = \begin{cases} \left(\dfrac{1}{\hat{\lambda}_i + t^{-1}}\right)\left(\hat{\lambda}_i x_{i,\text{meas}} + t^{-1} v_i\right), & x_{i,\text{meas}} \text{ measured} \\ v_i, & x_{i,\text{meas}} \text{ missing} \end{cases}. \tag{7.42}$$

In general, because the above penalties are separable in x, the separability property of proximal operators (7.10) can be exploited so that the proximal operator of *each* residual $x_i - x_{i,\text{meas}}$ can be determined independently. So, different penalty functions can be applied to each signal $w_i(k)$ in the original problem (7.30).

7.4.2 Structured artifact rejection

In the structured artifact rejection problem

$$\begin{array}{ll} \text{minimize} & \|W_{r,N}\|_* + \lambda \displaystyle\sum_{k=0}^{N-q} \|z(k)\|_2 \\ \text{subject to} & w^l - w^l_{\text{meas}} = Pz \\ & w_i(k) = w_{i,\text{meas}}(k), \quad i \neq l, k \in \mathcal{K}_i, \end{array} \tag{7.43}$$

the overlapping group sparsity penalty does not have a closed-form proximal operator [OJV11, YLY11]. The PDDR algorithm, which was used in Section 7.4.1, is most efficient when these proximal operators can be easily computed, and so is a poor choice here. The problem is instead solved using ADMM.

First, (7.43) is expressed in terms of a vector optimization variable x as

$$\begin{array}{ll} \text{minimize} & \|\text{mat}(Ax)\|_* + \lambda \displaystyle\sum_{k=0}^{N-q} \|z(k)\|_2 \\ \text{subject to} & C(x - x_{\text{meas}}) = Pz \\ & D(x - x_{\text{meas}}) = 0, \end{array} \tag{7.44}$$

where $C \in \mathbb{R}^{N \times Nd}$ is the binary matrix that selects the elements of w^l from x, and $D \in \mathbb{R}^{N_{\text{meas}} \times Nd}$ is the binary matrix that selects the $N_{\text{meas}} \leq N(d-1)$ measured elements from w_i, $i \neq l$. Next, dummy variables X and v are introduced that will facilitate the splitting of

the problem:

$$\begin{aligned} \text{minimize} \quad & \|X\|_* + \lambda \sum_{k=0}^{N-q} \|z(k)\|_2 \\ \text{subject to} \quad & C(x - x_{\text{meas}}) = Pv \\ & D(x - x_{\text{meas}}) = 0 \\ & v = z \\ & X = \text{mat}(Ax). \end{aligned} \tag{7.45}$$

This problem has the augmented Lagrangian

$$\begin{aligned} \mathcal{L}_\rho(x, z, X, v, y, u, \omega, \Gamma) = \|X\|_* + \lambda \sum_{k=0}^{N-q} \|z(k)\|_2 + y^\top (C(x - x_{\text{meas}}) - Pv) \\ + u^\top (D(x - x_{\text{meas}})) + \omega^\top (v - z) + \text{tr}\left[\Gamma^\top (X - \text{mat}(Ax))\right] \\ + \frac{\rho}{2}\|C(x - x_{\text{meas}}) - Pv\|_2^2 + \frac{\rho}{2}\|D(x - x_{\text{meas}})\|_2^2 \\ + \frac{\rho}{2}\|v - z\|_2^2 + \frac{\rho}{2}\|X - \text{mat}(Ax)\|_F^2. \end{aligned} \tag{7.46}$$

Because of the choice of dummy variables in (7.45), the Lagrangian is separable in (x, z) and (X, v). So, ADMM can be applied with the steps

$$\begin{aligned} (x^+, z^+) &= \underset{\hat{x}, \hat{z}}{\text{argmin}}\, \mathcal{L}_\rho(\hat{x}, \hat{z}, X, v, y, u, \omega, \Gamma) \\ (X^+, v^+) &= \underset{\hat{X}, \hat{v}}{\text{argmin}}\, \mathcal{L}_\rho(x^+, z^+, \hat{X}, \hat{v}, y, u, \omega, \Gamma) \\ y^+ &= y + \rho(C(x^+ - x_{\text{meas}}) - Pv^+) \\ u^+ &= u + \rho(D(x^+ - x_{\text{meas}})) \\ \omega^+ &= \omega + \rho(v^+ - z^+) \\ \Gamma^+ &= \Gamma + \rho(X^+ - \text{mat}(Ax^+)). \end{aligned} \tag{7.47}$$

The minimization steps have closed form solutions. They are:

$$x^+ = \left[C^\top C + D^\top D + A^\top A\right]^{-1} \left[C^\top C x_{\text{meas}} + C^\top P v + D^\top D x_{\text{meas}} + A^\top \text{vec}(X) + \rho^{-1}\left(a - C^\top y - D^\top u\right)\right],$$

$$z(k)^+ = \left(1 - \frac{\lambda}{\rho\|v + \rho^{-1}\omega\|_2}\right)_+ \cdot \left(v + \rho^{-1}\omega\right), \qquad k = 0, \ldots, N-q, \tag{7.48}$$

$$X^+ = \sum_{i=1}^{r} (\sigma_i - \rho^{-1})_+ u_i v_i^\top, \qquad \text{given the SVD } \sum_{i=1}^{r} \sigma_i u_i v_i^\top = \text{mat}(Ax) - \rho^{-1}\Gamma,$$

$$v^+ = z + \rho^{-1}\left(P^\top y - \omega\right),$$

where a is the vector constant such that

$$a^\top x = \text{tr}\left[\Gamma^\top \text{mat}(Ax)\right]. \tag{7.49}$$

7.5 System clustering

This section demonstrates how to solve the LTI system clustering method of Chapter 5, specifically, the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{L}\sum_{j=1}^{L} \|G_{ij}\|_F + \frac{\lambda}{2}\|\mathcal{O} - \mathcal{O}G\|_F^2 \\ \text{subject to} & G_{ii} = 0, \quad i = 1, \ldots, L, \end{array} \tag{7.50}$$

for G, where

$$\mathcal{O} = \begin{bmatrix} \mathcal{O}_1 & \cdots & \mathcal{O}_L \end{bmatrix}, \qquad G = \begin{bmatrix} G_{11} & \cdots & G_{L1} \\ \vdots & \ddots & \vdots \\ G_{1L} & \cdots & G_{LL} \end{bmatrix}. \tag{7.51}$$

7.5.1 Problem rearrangement

The problem (7.50) is rearranged in three steps. First, it is separated by noting that each block column of G can be computed independently. So, the problem can be rewritten as L

smaller subproblems of the form

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{L} \|X(j)\|_F + \frac{\lambda}{2}\|\mathcal{O}_i - \mathcal{O}X\|_F^2 \\ \text{subject to} \quad & X(i) = 0, \end{aligned} \tag{7.52}$$

for $i = 1, \ldots, L$, where X is the optimization variable that corresponds to

$$X = \begin{bmatrix} X(1) \\ \vdots \\ X(L) \end{bmatrix} = \begin{bmatrix} G_{i1} \\ \vdots \\ G_{iL} \end{bmatrix} \in \mathbb{R}^{nL \times n}. \tag{7.53}$$

Each of these problems can be solved in parallel, and the complete solution G reconstructed by simply aggregating the block columns of each X.

The second step is to vectorize the problem. Equation (7.52) can be rewritten with a vector variable x as

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{L} \|x(j)\|_2 + \frac{\lambda}{2}\|Ax - b\|_2^2 \\ \text{subject to} \quad & x(i) = 0 \end{aligned} \tag{7.54}$$

by choosing

$$x = \begin{bmatrix} x(1) \\ \vdots \\ x(L) \end{bmatrix} = \begin{bmatrix} \operatorname{vec}(X(1)) \\ \vdots \\ \operatorname{vec}(X(L)) \end{bmatrix} \in \mathbb{R}^{n^2 L}, \tag{7.55}$$

$$A = \begin{bmatrix} (I_n \otimes \mathcal{O}_1) & \cdots & (I_n \otimes \mathcal{O}_L) \end{bmatrix} \in \mathbb{R}^{npr \times n^2 L}, \tag{7.56}$$

$$b = \operatorname{vec}(\mathcal{O}_i) \in \mathbb{R}^{npr}. \tag{7.57}$$

The final step in the rearrangement is to incorporate the constraint in (7.54) into the objective function, as

$$\text{minimize} \quad \mathcal{I}_0(x(i)) + \sum_{\substack{j=1, \\ j \neq i}}^{L} \|x(j)\|_2 + \frac{\lambda}{2}\|Ax - b\|_2^2, \tag{7.58}$$

where $\mathcal{I}_0(\cdot)$ is again the convex indicator on the set $\{0\}$.

7.5.2 Proximal operators

The rearranged problem (7.58) is of the form

$$\text{minimize} \qquad f(x) + g(Ax), \tag{7.59}$$

where

$$\begin{aligned} f(x) &= \mathcal{I}_0(x(i)) + \sum_{\substack{j=1,\\ j\neq i}}^{L} \|x(j)\|_2, \\ g(y) &= \frac{\lambda}{2}\|y-b\|_2^2. \end{aligned} \tag{7.60}$$

The proximal operators for f and g^* have closed form solutions, and so the Chambolle–Pock algorithm is an ideal choice for solving this problem. The proximal operator of f can be found by noting that the function is separable in $x(i)$:

$$f(x) = \sum_{j=1}^{L} f_j(x(j)) \quad \Longrightarrow \quad \text{prox}_{tf}(v) = \begin{bmatrix} \text{prox}_{tf_1}(v(1)) \\ \vdots \\ \text{prox}_{tf_L}(v(L)) \end{bmatrix}. \tag{7.61}$$

Hence,

$$\begin{aligned} &\text{prox}_{tf_i}(v(i)) = 0, \\ &\text{prox}_{tf_j}(v(j)) = \begin{cases} 1 - t/\|v(j)\|_2, & \|v(j)\|_2 \geq t \\ 0, & \|v(j)\|_2 < t \end{cases}, \quad j \neq i. \end{aligned} \tag{7.62}$$

The proximal operator of g^* is

$$\text{prox}_{tg^*}(z) = \left(\frac{\lambda}{\lambda + t}\right)(z - tb). \tag{7.63}$$

These proximal operators, and choice of A, can then be directly applied to the Chambolle–Pock steps outlined in Section 7.3.3 in order to efficiently solve the problem.

7.6 Dictionary-based signal estimation

This final section provides details on how to solve the DBSE problem presented in Chapter 6:

$$\text{minimize} \qquad \|F(\hat{u}_{\text{meas}}, \hat{y})\|_* + \frac{\zeta}{2}\|\hat{y} - y_{\text{sim}}\|_2^2, \tag{7.64}$$

where $\hat{y} \in \mathbb{R}^N$ is the optimization variable. Using the results of Sections 7.4–7.5, this algorithm can be quickly evaluated. For simplicity, the following variation is solved

$$\text{minimize} \qquad \|F(\hat{u}, \hat{y})\|_* + \frac{\zeta}{2}\|\hat{y} - y_{\text{sim}}\|_2^2 + \sum_{i=1}^{m} \mathcal{I}_0(\hat{u}_i - \hat{u}_{i,\text{meas}}), \tag{7.65}$$

in which $\hat{u}$ is a variable.

First, by defining

$$x = \left[\begin{array}{c} x_u \\ \hline x_y \end{array}\right] = \left[\begin{array}{c} \hat{u}_1 \\ \vdots \\ \hat{u}_m \\ \hline \hat{y} \end{array}\right], \qquad a = \left[\begin{array}{c} \hat{u}_{1,\text{meas}} \\ \vdots \\ \hat{u}_{m,\text{meas}} \end{array}\right], \qquad b = y_{\text{sim}}, \tag{7.66}$$

the problem is expressed as

$$\text{minimize} \qquad \|\text{mat}(Ax)\|_* + \frac{\zeta}{2}\|x_y - b\|_2^2 + \mathcal{I}_0(x_u - a). \tag{7.67}$$

This problem has the form

$$\text{minimize} \qquad f(x) + g(Ax) \tag{7.68}$$

where

$$\begin{aligned} f(x) &= \frac{\zeta}{2}\|x_y - b\|_2^2 + \mathcal{I}_0(x_u - a), \\ g(y) &= \|\text{mat}(y)\|_*. \end{aligned} \tag{7.69}$$

The proximal operator for g^* is

$$\text{prox}_{tg^*}(z) = z - t \cdot \text{vec}\left[\sum_{i=1}^{r} \left(\sigma_i - t^{-1}\right)_+ u_i v_i^\top\right]. \tag{7.70}$$

The proximal operator for f is found using the separability property:

$$\text{prox}_{tf}(v) = \left[\begin{array}{c} \text{prox}_{tf_u}(v_u) \\ \text{prox}_{tf_y}(v_y) \end{array}\right], \tag{7.71}$$

where

$$\begin{aligned} \text{prox}_{tf_u}(v_u) &= a, \\ \text{prox}_{tf_y}(v_y) &= \left(\frac{1}{\zeta + t^{-1}}\right)\left(\zeta b + t^{-1} v_y\right). \end{aligned} \tag{7.72}$$

With the proximal operators for f and g^* known, (7.67) can be solved using the Chambolle–Pock algorithm. The steps in this algorithm are provided in Section 7.3.3.

CHAPTER 8

Conclusions

This dissertation has examined the problem of noninvasive intracranial pressure (NICP) estimation from the perspective of linear dynamical systems. By leveraging past work in subspace system identification and convex optimization, new and modular methods have been presented that add value to this larger research problem. These methods were framed using models and notation that are typical in linear systems theory, and so they apply to a large variety of applications beyond NICP estimation.

The first set of methods provide a way to find approximations of signals that exhibit dynamics of low complexity. Specifically, these signal approximations typically produce *low-order* linear dynamical systems. The main use cases of these methods are when signals are corrupted by artifacts, or contain missing data. Both of these scenarios are common for cerebral blood flow velocity (CBFV) signals, which are widely used in NICP estimators. The suite of methods was able to remove real CBFV artifacts and impute sections of missing data with accuracy.

The second set of methods can be used to cluster dynamical systems. In NICP estimation this process can be used to group patients by the similarity of their signal dynamics. These methods are important for mitigating the effect of large inter-patient variability: dynamical models can be constructed across a small, targeted group of patients, rather than across an entire patient population. The method was able to correctly cluster patient models with their labels removed, demonstrating that it is effective on models of ICP dynamics.

The final set of methods, called *dictionary-based signal estimation*, can be used to estimate missing signals of a dynamical system when a dictionary of similar systems is known. The process combines the incomplete information on the current system with the complete

information on these past systems. This method can be applied to NICP estimation, where the dictionary is of past patients who have had their ICP measured after an invasive procedure, and the current system is a patient whose ICP signal is missing.

The methods presented above are posed as convex optimization problems, which are well-understood and so provide some key advantages. For example, it is easy to extend these methods to new scenarios by introducing new penalty functions, constraints, or weights to their formulation. Also, because there are ongoing advances to improve the algorithms that compute the output of the methods, their efficiency will improve as the literature does.

There are two main directions for future work. The first is to continue to move this work towards future clinical application. Because the components of this dissertation have been constructed in a modular manner, they can be integrated into a variety of different existing NICP estimation frameworks, for example. In order to advance these approaches closer to true clinical applicability, a larger body of NICP data will also be necessary, including a variety of diagnoses, clinical settings, and geographic locations.

The second direction is to apply these data methods to new applications. The technical content of this dissertation has been written in the language of linear dynamical systems, which are widely used, from engineering to the social sciences. So, the methods of this dissertation can be readily applied to a variety of problems, and will hopefully provide new and useful steps forward across different fields.

REFERENCES

[Aas12] Rune Aaslid. *Transcranial doppler sonography*. Springer Science & Business Media, 2012.

[ABK08] Shushma Aggarwal, David M Brooks, Yoogoo Kang, Peter K Linden, and John F Patzer. "Noninvasive monitoring of cerebral perfusion pressure in patients with acute liver failure using transcranial doppler ultrasonography." *Liver Transplantation*, **14**(7):1048–1057, 2008.

[AHT10] Yaseen M Arabi, Samir Haddad, Hani M Tamim, Abdulaziz Al-Dawood, Saad Al-Qahtani, Ahmad Ferayan, Ibrahim Al-Abdulmughni, Jalal Al-Oweis, and Asia Rugaan. "Mortality reduction after implementing a clinical practice guidelines–based management protocol for severe traumatic brain injury." *Journal of critical care*, **25**(2):190–195, 2010.

[AKG86] Paula J Aucoin, Helen Rosen Kotilainen, Nelson M Gantz, Robin Davidson, Peter Kellogg, and Bernard Stone. "Intracranial pressure monitors. Epidemiologic study of risk factors and infections." *The American journal of medicine*, **80**(3):369–376, 1986.

[ALL86] R Aaslid, T Lundar, KF Lindegaard, and H Nornes. "Estimation of cerebral perfusion pressure from arterial blood pressure and transcranial Doppler recordings." In *Intracranial pressure VI*, pp. 226–229. Springer, 1986.

[AS12] Mustafa Ayazoglu and Mario Sznaier. "An algorithm for fast constrained nuclear norm minimization and applications to systems identification." In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 3469–3475. IEEE, 2012.

[AYS13] Mustafa Ayazoglu, Burak Yilmaz, Mario Sznaier, and Octavia Camps. "Finding causal interactions in video sequences." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3575–3582, 2013.

[BBS10] Giovanna Brandi, Markus Béchir, Susanne Sailer, Christoph Haberthür, Reto Stocker, and John F Stover. "Transcranial color-coded duplex sonography allows to assess cerebral perfusion pressure noninvasively following severe traumatic brain injury." *Acta neurochirurgica*, **152**(6):965–972, 2010.

[BF10] José M Bioucas-Dias and Mário AT Figueiredo. "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing." In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pp. 1–4. IEEE, 2010.

[Bir14] Sarah Lynn Bird. *Optimizing Resource Allocations for Dynamic Interactive Applications*. University of California, Berkeley, 2014.

[BJM12] Francis Bach, Rodolphe Jenatton, Julien Mairal, Guillaume Obozinski, et al. "Optimization with sparsity-inducing penalties." *Foundations and Trends in Machine Learning*, **4**(1):1–106, 2012.

[BLA10] Anders Behrens, Niklas Lenfeldt, Khalid Ambarki, Jan Malm, Anders Eklund, and Lars-Owe Koskinen. "Transcranial Doppler pulsatility index: not an accurate method to assess intracranial pressure." *Neurosurgery*, **66**(6):1050–1057, 2010.

[BMW77] Donald P Becker, J Douglas Miller, John D Ward, Richard P Greenberg, Harold F Young, and Romas Sakalas. "The outcome from severe head injury with early diagnosis and intensive management." *Journal of neurosurgery*, **47**(4):491–502, 1977.

[BRR04] Johan Bellner, Bertil Romner, Peter Reinstrup, Karl-Axel Kristiansson, Erik Ryding, and Lennart Brandt. "Transcranial Doppler sonography pulsatility index (PI) reflects intracranial pressure (ICP)." *Surgical neurology*, **62**(1):45–51, 2004.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[BV18] Stephen Boyd and Lieven Vandenberghe. *Introduction to Applied Linear Algebra—Vectors, Matrices, and Least Squares*. Cambridge university press, 2018.

[CB17] Niladri Chatterji and Peter L Bartlett. "Alternating minimization for dictionary learning with random initialization." In *Advances in Neural Information Processing Systems*, pp. 1994–2003, 2017.

[CLM11] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. "Robust principal component analysis?" *Journal of the ACM (JACM)*, **58**(3):11, 2011.

[CMD92] Kwan-Hon Chan, J Douglas Miller, N Mark Dearden, Peter JD Andrews, and Susan Midgley. "The effect of changes in cerebral perfusion pressure upon middle cerebral artery blood flow velocity and jugular bulb venous oxygen saturation after severe brain injury." *Journal of neurosurgery*, **77**(1):55–61, 1992.

[CMK93] Randall M Chesnut, Lawrence F Marshall, Melville R Klauber, Barbara A Blunt, Nevan Baldwin, Howard M Eisenberg, John A Jane, Anthony Marmarou, and Mary A Foulkes. "The role of secondary brain injury in determining outcome from severe head injury." *The Journal of trauma*, **34**(2):216–222, 1993.

[CMS98] Marek Czosnyka, Basil F Matta, Piotr Smielewski, Peter J Kirkpatrick, and John D Pickard. "Cerebral perfusion pressure in head-injured patients: a noninvasive assessment using transcranial Doppler ultrasonography." *Journal of neurosurgery*, **88**(5):802–808, 1998.

[CMS14] Emmanuel Candès, Lester Mackey, and Mahdi Soltanolkotabi. "From subspace clustering to full-rank matrix completion." In *Unpublished abstract*, 2014.

[CP04] Marek Czosnyka and John D Pickard. "Monitoring and interpretation of intracranial pressure." *Journal of Neurology, Neurosurgery & Psychiatry*, **75**(6):813–821, 2004.

[CP10] Emmanuel J Candès and Yaniv Plan. "Matrix completion with noise." *Proceedings of the IEEE*, **98**(6):925–936, 2010.

[CP11] Antonin Chambolle and Thomas Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging." *Journal of Mathematical Imaging and Vision*, **40**(1):120–145, 2011.

[CTO17] Nancy Carney, Annette M Totten, Cindy O'reilly, Jamie S Ullman, Gregory WJ Hawryluk, Michael J Bell, Susan L Bratton, Randall Chesnut, Odette A Harris, Niranjan Kissoon, et al. "Guidelines for the management of severe traumatic brain injury." *Neurosurgery*, **80**(1):6–15, 2017.

[DD00] Katrien De Cock and Bart De Moor. "Subspace angles and distances between ARMA models." In *Proc. of the Intl. Symp. of Math. Theory of networks and systems*, volume 1. Citeseer, 2000.

[DD02] Katrien De Cock and Bart De Moor. "Subspace angles between ARMA models." *Systems & Control Letters*, **46**(4):265–270, 2002.

[DDD97] Bart De Moor, Peter De Gersem, Bart De Schutter, and Wouter Favoreel. "DAISY: A database for identification of systems." *Journal A*, **38**:4–5, 1997.

[DHS12] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[DJS12] Mahua Dey, Jennifer Jaffe, Agnieszka Stadnik, and Issam A Awad. "External ventricular drainage for intraventricular hemorrhage." *Current neurology and neuroscience reports*, **12**(1):24–33, 2012.

[DNZ13] Elena Daskalaki, Kirsten Nørgaard, Thomas Züger, Aikaterini Prountzou, Peter Diem, and Stavroula Mougiakakou. "An early warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models." *Journal of diabetes science and technology*, **7**(3):689–698, 2013.

[Don95] David L Donoho. "De-noising by soft-thresholding." *IEEE transactions on information theory*, **41**(3):613–627, 1995.

[DR56] Jim Douglas and Henry H Rachford. "On the numerical solution of heat conduction problems in two and three space variables." *Transactions of the American mathematical Society*, **82**(2):421–439, 1956.

[EB92] Jonathan Eckstein and Dimitri P Bertsekas. "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators." *Mathematical Programming*, **55**(1):293–318, 1992.

[Efr94] Bradley Efron. "Missing data, imputation, and the bootstrap." *Journal of the American Statistical Association*, **89**(426):463–475, 1994.

[EV09] Ehsan Elhamifar and René Vidal. "Sparse subspace clustering." In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 2790–2797. IEEE, 2009.

[EV13] Ehsan Elhamifar and Rene Vidal. "Sparse subspace clustering: Algorithm, theory, and applications." *IEEE transactions on pattern analysis and machine intelligence*, **35**(11):2765–2781, 2013.

[Faz02] Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[FGC12] Arash Farahvar, Linda M Gerber, Ya-Lin Chiu, Nancy Carney, Roger Härtl, and Jamshid Ghajar. "Increased mortality in patients with severe traumatic brain injury treated without intracranial pressure monitoring." *Journal of neurosurgery*, **117**(4):729–734, 2012.

[FHB01] Maryam Fazel, Haitham Hindi, and Stephen P Boyd. "A rank minimization heuristic with application to minimum order system approximation." In *American Control Conference, 2001. Proceedings of the 2001*, volume 6, pp. 4734–4739. IEEE, 2001.

[FHP95] Karl J Friston, Andrew P Holmes, JB Poline, PJ Grasby, SCR Williams, Richard SJ Frackowiak, and Robert Turner. "Analysis of fMRI time-series revisited." *Neuroimage*, **2**(1):45–53, 1995.

[FPE94] Gene F Franklin, J David Powell, Abbas Emami-Naeini, and J David Powell. *Feedback control of dynamic systems*, volume 3. Addison-Wesley Reading, MA, 1994.

[FR91] Thomas MJ Fruchterman and Edward M Reingold. "Graph drawing by force-directed placement." *Software: Practice and experience*, **21**(11):1129–1164, 1991.

[FSP98] Mads C Forchhammer, N Chr Stenseth, Eric Post, and R Landvatn. "Population dynamics of Norwegian red deer: density–dependence and climatic variation." *Proceedings of the Royal Society of London B: Biological Sciences*, **265**(1393):341–350, 1998.

[FTW04] Samir M Fakhry, Arthur L Trask, Maureen A Waller, Dorraine D Watts, et al. "Management of brain-injured patients by an evidence-based medicine protocol improves outcomes and decreases hospital charges." *Journal of Trauma and Acute Care Surgery*, **56**(3):492–500, 2004.

[GB] Michael Grant and Stephen Boyd. "CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 Beta, Sep. 2013." *Available on-line at http://cvxr.com/cvx*.

[GDD98] LL Guyot, C Dowling, FG Diaz, and Daniel B Michael. *Cerebral monitoring devices: analysis of complications*. Springer, 1998.

[GM76] Daniel Gabay and Bertrand Mercier. "A dual algorithm for the solution of nonlinear variational problems via finite element approximation." *Computers & Mathematics with Applications*, **2**(1):17–40, 1976.

[GR15] Christian Grussler and Anders Rantzer. "On optimal low-rank approximation of non-negative matrices." In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, pp. 5278–5283. IEEE, 2015.

[GRG18] Christian Grussler, Anders Rantzer, and Pontus Giselsson. "Low-rank optimization with convex constraints." *IEEE Transactions on Automatic Control*, 2018.

[GTS15] Euhanna Ghadimi, André Teixeira, Iman Shames, and Mikael Johansson. "Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems." *IEEE Transactions on Automatic Control*, **60**(3):644–658, 2015.

[HA12] Samir H Haddad and Yaseen M Arabi. "Critical care management of severe traumatic brain injury in adults." *Scandinavian journal of trauma, resuscitation and emergency medicine*, **20**(1):1, 2012.

[Hal15] John E Hall. *Guyton and Hall Textbook of Medical Physiology*. Elsevier Health Sciences, 2015.

[Har75] John A Hartigan. "Clustering algorithms." 1975.

[HNB06] Xiao Hu, Valeriy Nenov, Marvin Bergsneider, and Neil Martin. "A data mining framework of noninvasive intracranial pressure assessment." *Biomedical Signal Processing and Control*, **1**(1):64–77, 2006.

[HNG06] Xiao Hu, Valeriy Nenov, Thomas C Glenn, Luzius A Steiner, Marek Czosnyka, Marvin Bergsneider, and Neil Martin. "Nonlinear analysis of cerebral hemodynamic and intracranial pressure signals for characterization of autoregulation." *Biomedical Engineering, IEEE Transactions on*, **53**(2):195–209, 2006.

[HSS08] Aric Hagberg, Pieter Swart, and Daniel S Chult. "Exploring network structure, dynamics, and function using NetworkX." Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[Hub64] Peter J Huber et al. "Robust estimation of a location parameter." *The Annals of Mathematical Statistics*, **35**(1):73–101, 1964.

[HW79] John A Hartigan and Manchek A Wong. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**(1):100–108, 1979.

[HXS09] Xiao Hu, Peng Xu, Fabien Scalzo, Paul Vespa, and Marvin Bergsneider. "Morphological clustering and analysis of continuous intracranial pressure." *IEEE Transactions on Biomedical Engineering*, **56**(3):696–705, 2009.

[HXW10] Xiao Hu, Peng Xu, Shaozhi Wu, Shadnaz Asgari, and Marvin Bergsneider. "A data mining framework for time series estimation." *Journal of biomedical informatics*, **43**(2):190–199, 2010.

[JHL98] Tzyy-Ping Jung, Colin Humphries, Te-Won Lee, Scott Makeig, Martin J McKeown, Vicente Iragui, Terrence J Sejnowski, et al. "Extended ICA removes artifacts from electroencephalographic recordings." *Advances in neural information processing systems*, pp. 894–900, 1998.

[Joh67] Stephen C Johnson. "Hierarchical clustering schemes." *Psychometrika*, **32**(3):241–254, 1967.

[Jon97] Bradley Jones. "MATLAB Statistics Toolbox." 1997.

[JW98] Magnus Jansson and Bo Wahlberg. "On consistency of subspace methods for system identification." *Automatica*, **34**(12):1507–1519, 1998.

[KA86] Robert Kohn and Craig F Ansley. "Estimation, prediction, and interpolation for ARIMA models with missing data." *Journal of the American Statistical Association*, **81**(395):751–761, 1986.

[KB14] Stefan Kersting and Martin Buss. "Metrics to compare and control dynamical systems." In *Control Conference (ECC), 2014 European*, pp. 738–743. IEEE, 2014.

[KCB88] J Klingelhöfer, B Conrad, R Benecke, D Sander, and E Markakis. "Evaluation of intracranial pressure from transcranial Doppler studies in cerebral disease." *Journal of neurology*, **235**(3):159–162, 1988.

[KCS98] Terry Bo-Jau Kuo, Chang-Ming Chern, Wen-Yung Sheng, Wen-Jang Wong, and Han-Hwa Hu. "Frequency domain analysis of cerebral blood flow velocity and its correlation with arterial blood pressure." *Journal of Cerebral Blood Flow & Metabolism*, **18**(3):311–318, 1998.

[KHJ01] Gary King, James Honaker, Anne Joseph, and Kenneth Scheve. "Analyzing incomplete political science data: An alternative algorithm for multiple imputation." In *American Political Science Association*, volume 95, pp. 49–69. Cambridge Univ Press, 2001.

[KHP13] Sunghan Kim, Robert Hamilton, Stacy Pineles, Marvin Bergsneider, and Xiao Hu. "Noninvasive intracranial hypertension detection utilizing semisupervised learning." *IEEE Transactions on Biomedical Engineering*, **60**(4):1126–1133, 2013.

[KHV08] FM Kashif, T Heldt, and GC Verghese. "Model-based estimation of intracranial pressure and cerebrovascular autoregulation." In *Computers in Cardiology, 2008*, pp. 369–372. IEEE, 2008.

[KL13] Nurhan Karaboga and Fatma Latifoglu. "Elimination of noise on transcranial Doppler signal using IIR filters designed with artificial bee colony–ABC-algorithm." *Digital Signal Processing*, **23**(3):1051–1058, 2013.

[KSB12] Sunghan Kim, Fabien Scalzo, Marvin Bergsneider, Paul Vespa, Neil Martin, and Xiao Hu. "Noninvasive intracranial pressure assessment based on a data-mining approach using a nonlinear mapping function." *Biomedical Engineering, IEEE Transactions on*, **59**(3):619–626, 2012.

[KSM08] Heidi Harbison Kimberly, Sachita Shah, Keith Marill, and Vicki Noble. "Correlation of optic nerve sheath diameter with direct measurement of intracranial pressure." *Academic Emergency Medicine*, **15**(2):201–204, 2008.

[KT02] Alan Kirman and Gilles Teyssiere. "Microeconomic models for long memory in the volatility of financial time series." *Studies in Nonlinear Dynamics & Econometrics*, **5**(4), 2002.

[KVN12] Faisal M Kashif, George C Verghese, Vera Novak, Marek Czosnyka, and Thomas Heldt. "Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure." *Science translational medicine*, **4**(129):129ra44–129ra44, 2012.

[KWP85] Robert K Kanter, Leonard B Weiner, Anne Marie Patti, and Linda K Robson. "Infectious complications and duration of intracranial pressure monitoring." *Critical care medicine*, **13**(10):837–839, 1985.

[KX10] Seyoung Kim and Eric P Xing. "Tree-guided group lasso for multi-task regression with structured sparsity." 2010.

[Lau05] Alan J Laub. *Matrix analysis for scientists and engineers*. Siam, 2005.

[LBP09] Daniel TH Lai, Rezaul K Begg, and Marimuthu Palaniswami. "Computational intelligence in gait research: a perspective on current applications and future challenges." *IEEE Transactions on Information Technology in Biomedicine*, **13**(5):687–702, 2009.

[LC11] Jinseok Lee and Ki H Chon. "Time-varying autoregressive model-based multiple modes particle filtering algorithm for respiratory rate extraction from pulse oximeter." *IEEE Transactions on Biomedical Engineering*, **58**(3):790–794, 2011.

[Lev77] Allan B Levin. "The use of a fiberoptic intracranial pressure monitor in clinical practice." *Neurosurgery*, **1**(3):266–271, 1977.

[LGB10] Yenny Leal, Winston Garcia-Gabin, Jorge Bondia, Eduardo Esteve, Wifredo Ricart, Jose-Manuel Fernández-Real, and Josep Vehí. "Real-time glucose estimation algorithm for continuous glucose monitoring using autoregressive models." *Journal of diabetes science and technology*, **4**(2):391–403, 2010.

[LHV13] Zhang Liu, Anders Hansson, and Lieven Vandenberghe. "Nuclear norm system identification with missing inputs and outputs." *Systems & Control Letters*, **62**(8):605–612, 2013.

[Lju98] Lennart Ljung. *System identification*. Springer, 1998.

[Lju07] Lennart Ljung. "System Identification Toolbox for Use with MATLAB." 2007.

[LL95] WK Li and K Lam. "Modelling asymmetry in stock returns by a threshold autoregressive conditional heteroscedastic model." *The Statistician*, pp. 333–341, 1995.

[LM79] Pierre-Louis Lions and Bertrand Mercier. "Splitting algorithms for the sum of two nonlinear operators." *SIAM Journal on Numerical Analysis*, **16**(6):964–979, 1979.

[LSK00] Fin Stolze Larsen, Gitte Strauss, Gitte Moos Knudsen, Tina Maria Herzog, Bent Adel Hansen, and Niels H Secher. "Cerebral perfusion, cardiac output, and arterial pressure in patients with fulminant hepatic failure." *Critical care medicine*, **28**(4):996–1000, 2000.

[LST13] H Lehman Li-wei, Mohammed Saeed, Daniel Talmor, Roger Mark, and Atul Malhotra. "Methods of blood pressure measurement in the ICU." *Critical care medicine*, **41**(1):34, 2013.

[LV09] Zhang Liu and Lieven Vandenberghe. "Interior-point method for nuclear norm approximation with application to system identification." *SIAM Journal on Matrix Analysis and Applications*, **31**(3):1235–1256, 2009.

[MAL84] C Glen Mayhall, Nancy H Archer, V Archer Lamb, Alice C Spadora, Jane W Baggett, John D Ward, and Raj K Narayan. "Ventriculostomy-related infections." *New England Journal of Medicine*, **310**(9):553–559, 1984.

[Mar00] Richard J Martin. "A metric for ARMA processes." *IEEE transactions on Signal Processing*, **48**(4):1164–1170, 2000.

[MBG81] J Douglas Miller, John F Butterworth, Steven K Gudeman, J Edward Faulkner, Sung C Choi, John B Selhorst, John W Harbison, Harry A Lutz, Harold F Young, and Donald P Becker. "Further experience in the management of severe head injury." *Journal of neurosurgery*, **54**(3):289–299, 1981.

[MDV89] Marc Moonen, Bart De Moor, Lieven Vandenberghe, and Joos Vandewalle. "On- and off-line identification of linear state-space models." *International Journal of Control*, **49**(1):219–232, 1989.

[Med15] Association for the Advancement of Medical Instrumentation. "ANSI/AAMI NS28/Ed.1, Intracranial pressure monitoring devices." 2015.

[MF10] Karthik Mohan and Maryam Fazel. "Iterative reweighted least squares for matrix rank minimization." In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pp. 653–661. IEEE, 2010.

[MFH16] M. Matthews, A. Fanelli, and T Heldt. "An Embedded Device for Real-Time Noninvasive ICP Estimation." In *Intracranial Pressure and Neuromonitoring, 16th International Symposium on*, 2016.

[MMK08] Pierre D Mourad, Brandt Mohr, Michel Kliot, Robert CA Frederickson, R Lee Thompson, and L Jason. "Systems and methods for determining intracranial pressure non-invasively and acoustic transducer assemblies for use in such systems.", December 1 2008. US Patent App. 12/325,977.

[MMS08] I Marshall, I MacCormick, R Sellar, and I Whittle. "Assessment of factors affecting MRI measurement of intracranial volume changes and elastance index." *British journal of neurosurgery*, **22**(3):389–397, 2008.

[MMW87] Anthony Marmarou, Angelo L Maset, John D Ward, Sung Choi, Danny Brooks, Harry A Lutz, Richard J Moulton, J Paul Muizelaar, Antonio Desalles, and Harold F Young. "Contribution of CSF and vascular factors to elevation of ICP in severely head-injured patients." *Journal of neurosurgery*, **66**(6):883–890, 1987.

[Mor65] Jean-Jacques Moreau. "Proximité et dualité dans un espace hilbertien." *Bull. Soc. Math. France*, **93**(2):273–299, 1965.

[Mos10] APS Mosek. "The MOSEK optimization software." *Online at http://www. mosek. com*, **54**:2–1, 2010.

[MV90] Marc Moonen and Joos Vandewalle. "QSVD approach to on-and off-line state-space identification." *International Journal of Control*, **51**(5):1133–1146, 1990.

[MWS98] E Münch, R Weigel, P Schmiedek, and L Schürer. "The Camino intracranial pressure device in clinical practice: reliability, handling characteristics and complications." *Acta neurochirurgica*, **140**(11):1113–1120, 1998.

[MWV06] Ivan Markovsky, Jan C Willems, Sabine Van Huffel, and Bart De Moor. *Exact and approximate modeling of linear systems: A behavioral approach*, volume 11. SIAM, 2006.

[Neu00] Trauma.org Neurotrauma. "Cerebral Perfusion Pressure." 2000.

[Neu11] American Association of Neuroscience Nurses et al. "Care of the patient undergoing intracranial pressure monitoring/external ventricular drainage or lumbar drainage." *Glenview (IL): American Association of Neuroscience Nurses*, pp. 1–38, 2011.

[NJW02] Andrew Y Ng, Michael I Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm." In *Advances in neural information processing systems*, pp. 849–856, 2002.

[NKB82] Raj K Narayan, Pulla RS Kishore, Donald P Becker, John D Ward, Gregory G Enas, Richard P Greenberg, A Domingues Da Silva, Maurice H Lipper, Sung C Choi, C Glen Mayhall, et al. "Intracranial pressure: to monitor or not to monitor? A review of our experience with severe head injury." *Journal of neurosurgery*, **56**(5):650–659, 1982.

[NLR15] Robert Nishihara, Laurent Lessard, Benjamin Recht, Andrew Packard, and Michael I Jordan. "A general analysis of the convergence of ADMM." *arXiv preprint arXiv:1502.02009*, 2015.

[NWR10] Hugh Nolan, Robert Whelan, and RB Reilly. "FASTER: fully automated statistical thresholding for EEG artifact rejection." *Journal of neuroscience methods*, **192**(1):152–162, 2010.

[OGS12] J Olesen, A Gustavsson, Mikael Svensson, H-U Wittchen, and B Jönsson. "The economic cost of brain disorders in Europe." *European Journal of Neurology*, **19**(1):155–162, 2012.

[OJV11] Guillaume Obozinski, Laurent Jacob, and Jean-Philippe Vert. "Group lasso with overlaps: the latent group lasso approach." *arXiv preprint arXiv:1110.0413*, 2011.

[OV14] Daniel O'Connor and Lieven Vandenberghe. "Primal-dual decomposition by operator splitting and applications to image deblurring." *SIAM Journal on Imaging Sciences*, **7**(3):1724–1754, 2014.

[OY02] Katsuhiko Ogata and Yanjuan Yang. *Modern control engineering*, volume 4. Prentice hall India, 2002.

[PB13] Neal Parikh and Stephen Boyd. "Proximal algorithms." *Foundations and Trends in optimization*, **1**(3):123–231, 2013.

[Pel15] Kristiaan Pelckmans. "Lecture notes for a course on system identification." 2015.

[PF11] Gabriel Peyré and Jalal Fadili. "Group sparsity with overlapping partition functions." In *Signal Processing Conference, 2011 19th European*, pp. 303–307. IEEE, 2011.

[PGK04] Paul Park, Hugh JL Garton, Mary Jo Kocan, and B Gregory Thompson. "Risk of infection with prolonged ventricular catheterization." *Neurosurgery*, **55**(3):594–601, 2004.

[PT94] CG Paramore and DA Turner. "Relative risks of ventriculostomy infection and morbidity." *Acta neurochirurgica*, **127**(1-2):79–84, 1994.

[RAS03] Patricia B Raksin, Noam Alperin, Anusha Sivaramakrishnan, Sushma Surapaneni, and Terry Lichtor. "Noninvasive intracranial compliance and pressure based on dynamic magnetic resonance imaging of blood flow and cerebrospinal fluid flow: review of principles, implementation, and other noninvasive approaches." *Neurosurgical focus*, **14**(4):1–8, 2003.

[RB16] Ernest K Ryu and Stephen Boyd. "Primer on monotone operator methods." *Appl. Comput. Math*, **15**(1):3–43, 2016.

[RBA12] PH Raboel, J Bartek, M Andresen, BM Bellander, and B Romner. "Intracranial pressure monitoring: invasive versus non-invasive methods—a review." *Critical care research and practice*, **2012**, 2012.

[RBN06] Deborah J Rubens, Shweta Bhatt, Shannon Nedelka, and Jeanne Cullinan. "Doppler artifacts and pitfalls." *Ultrasound Clinics*, **1**(1):79–109, 2006.

[RCV09] Avinash Ravichandran, Rizwan Chaudhry, and René Vidal. "View-invariant dynamic texture recognition using a bag of dynamical systems." In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1651–1657. IEEE, 2009.

[RCV13] Avinash Ravichandran, Rizwan Chaudhry, and Rene Vidal. "Categorizing dynamic textures using a bag of dynamical systems." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(2):342–353, 2013.

[RFP10] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization." *SIAM review*, **52**(3):471–501, 2010.

[RL05] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.

[RLT06] Wesley Rutland-Brown, Jean A Langlois, Karen E Thomas, and Yongli Lily Xi. "Incidence of traumatic brain injury in the United States, 2003." *The Journal of head trauma rehabilitation*, **21**(6):544–548, 2006.

[RMN18] Saiprasad Ravishankar, Anna Ma, and Deanna Needell. "Analysis of Fast Alternating Minimization for Structured Dictionary Learning." *arXiv preprint arXiv:1802.00518*, 2018.

[Roc76] R Tyrrell Rockafellar. "Monotone operators and the proximal point algorithm." *SIAM journal on control and optimization*, **14**(5):877–898, 1976.

[ROF92] Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms." *Physica D: Nonlinear Phenomena*, **60**(1-4):259–268, 1992.

[RP13] Bernardino Romera-Paredes and Massimiliano Pontil. "A new convex relaxation for tensor completion." In *Advances in Neural Information Processing Systems*, pp. 2967–2975, 2013.

[RRJ95] Michael J Rosner, Sheila D Rosner, and Alice H Johnson. "Cerebral perfusion pressure: management protocol and clinical results." *Journal of neurosurgery*, **83**(6):949–962, 1995.

[RSC79] Allen K Ream, Gerald D Silverberg, Steven D Corbin, Eugene V Schmidt, and Thomas B Fryer. "Epidural measurement of intracranial pressure." *Neurosurgery*, **5**(1):36–43, 1979.

[SAI14] Mario Sznaier, Mustafa Ayazoglu, and Tamer Inanc. "Fast structured nuclear norm minimization with applications to set membership systems identification." *IEEE Transactions on Automatic Control*, **59**(10):2837–2842, 2014.

[SAJ14] Hanie Sedghi, Anima Anandkumar, and Edmond Jonckheere. "Multi-step stochastic ADMM in high dimensions: Applications to sparse optimization and matrix decomposition." In *Advances in neural information processing systems*, pp. 2771–2779, 2014.

[SBH02] RW Sherman, RA Bowie, MME Henfrey, RP Mahajan, and D Bogod. "Cerebral haemodynamics in pregnancy and pre-eclampsia as assessed by transcranial Doppler ultrasonography." *British journal of anaesthesia*, **89**(5):687–692, 2002.

[SC04] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[SC12] Mahdi Soltanolkotabi, Emmanuel J Candès, et al. "A geometric analysis of subspace clustering with outliers." *The Annals of Statistics*, **40**(4):2195–2238, 2012.

[SCR03] Bernhard Schmidt, Marek Czosnyka, Andreas Raabe, Hilal Yahya, Jens Jürgen Schwarze, Dieter Sackerer, Dirk Sander, and Jürgen Klingelhöfer. "Adaptive noninvasive assessment of intracranial pressure and cerebral autoregulation." *Stroke*, **34**(1):84–89, 2003.

[SCS00] Bernhard Schmidt, Marek Czosnyka, Jens Jürgen Schwarze, Dirk Sander, Werner Gerstner, Christianto Benjamin Lumenta, and Jürgen Klingelhöfer. "Evaluation of a method for noninvasive intracranial pressure assessment during infusion studies in patients with hydrocephalus." *Journal of neurosurgery*, **92**(5):793–800, 2000.

[SEC14] Mahdi Soltanolkotabi, Ehsan Elhamifar, Emmanuel J Candès, et al. "Robust subspace clustering." *The Annals of Statistics*, **42**(2):669–699, 2014.

[SKS97] Bernhard Schmidt, Jürgen Klingelhöfer, Jens Jürgen Schwarze, Dirk Sander, and Ingo Wittich. "Noninvasive prediction of intracranial pressure curves using transcranial Doppler ultrasonography and blood pressure curves." *Stroke*, **28**(12):2465–2472, 1997.

[SLY14] Wei Shi, Qing Ling, Kun Yuan, Gang Wu, and Wotao Yin. "On the Linear Convergence of the ADMM in Decentralized Consensus Optimization." *IEEE Trans. Signal Processing*, **62**(7):1750–1761, 2014.

[SM00] Jianbo Shi and Jitendra Malik. "Normalized cuts and image segmentation." *IEEE Transactions on pattern analysis and machine intelligence*, **22**(8):888–905, 2000.

[SM14] Nino Stocchetti and Andrew IR Maas. "Traumatic intracranial hypertension." *New England Journal of Medicine*, **370**(22):2121–2130, 2014.

[Smi12] Roy S Smith. "Nuclear norm minimization methods for frequency domain subspace identification." In *American Control Conference (ACC), 2012*, pp. 2689–2694. IEEE, 2012.

[SNS88] GÖRan Sundbärg, Carl-Henrik Nordström, and Sven Söderström. "Complications due to prolonged ventricular fluid pressure recording." *British journal of neurosurgery*, **2**(4):485–495, 1988.

[SPH08] Mihailo Stojnic, Farzad Parvaresh, and Babak Hassibi. "On the reconstruction of block-sparse signals with an optimal number of measurements." *arXiv preprint arXiv:0804.0041*, 2008.

[SSF04] Bruno Sinopoli, Luca Schenato, Massimo Franceschetti, Kameshwar Poolla, Michael I Jordan, and Shankar S Sastry. "Kalman filtering with intermittent observations." *IEEE transactions on Automatic Control*, **49**(9):1453–1464, 2004.

[Ste02] Karsten Sternickel. "Automatic pattern recognition in ECG time series." *Computer methods and programs in biomedicine*, **68**(2):109–115, 2002.

[SWS16] B. Schmidt, M. Weinhold, S. Streif, M. Czosnyka, and J. Klingelhofer. "Comparison of different calibration methods in a non-invasive ICP assessment model." In *Intracranial Pressure and Neuromonitoring, 16th International Symposium on*, 2016.

[TA07] Raghida Traboulsi and Paul Avan. "Transmission of infrasonic pressure waves from cerebrospinal to intralabyrinthine fluids through the human cochlear aqueduct: non-invasive measurements with otoacoustic emissions." *Hearing research*, **233**(1):30–39, 2007.

[TJ74] Graham Teasdale and Bryan Jennett. "Assessment of coma and impaired consciousness: a practical scale." *The Lancet*, **304**(7872):81–84, 1974.

[TKI86] T Takemae, Y Kosugi, J Ikebe, and Y Kumagai. "Electrical circuit model of cerebral circulation." *Iyo denshi to seitai kogaku. Japanese journal of medical electronics and biological engineering*, **24**(5):327, 1986.

[UIS95] Mauro Ursino, Maurizio Iezzi, and Nino Stocchetti. "Intracranial pressure dynamics in patients with acute brain damage: a critical analysis with the aid of a mathematical model." *IEEE Transactions on Biomedical Engineering*, **42**(6):529–540, 1995.

[UL97] Mauro Ursino and Carlo Alberto Lodi. "A simple mathematical model of the interaction between intracranial pressure and cerebral hemodynamics." *Journal of Applied Physiology*, **82**(4):1256–1269, 1997.

[Urs88a] Mauro Ursino. "A mathematical study of human intracranial hydrodynamics part 1—the cerebrospinal fluid pulse pressure." *Annals of biomedical engineering*, **16**(4):379–401, 1988.

[Urs88b] Mauro Ursino. "A mathematical study of human intracranial hydrodynamics part 2—Simulation of clinical tests." *Annals of biomedical engineering*, **16**(4):403–416, 1988.

[USK04] AW Unterberg, J Stover, B Kress, and KL Kiening. "Edema and brain trauma." *Neuroscience*, **129**(4):1019–1027, 2004.

[Van18] Lieven Vandenberghe. *ECE133A—Applied Numerical Computing (lecture notes)*. 2018.

[VBY73] John K Vries, Donald P Becker, and Harold F Young. "A subarachnoid screw for monitoring intracranial pressure: Technical note." *Journal of neurosurgery*, **39**(3):416–419, 1973.

[VD94] Peter Van Overschee and Bart De Moor. "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems." *Automatica*, **30**(1):75–93, 1994.

[VD12] Peter Van Overschee and BL De Moor. *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media, 2012.

[VH14] Michel Verhaegen and Anders Hansson. "Nuclear norm subspace identification (N2SID) for short data batches." *IFAC Proceedings Volumes*, **47**(3):9528–9533, 2014.

[VH15] Michel Verhaegen and Anders Hansson. "N2SID: Nuclear norm subsystem identification." *arXiv preprint arXiv:1501.04495*, 2015.

[VH16] Michel Verhaegen and Anders Hansson. "N2SID: Nuclear norm subspace identification of innovation models." *Automatica*, **72**:57–63, 2016.

[Von07] Ulrike Von Luxburg. "A tutorial on spectral clustering." *Statistics and computing*, **17**(4):395–416, 2007.

[VSR96] JH Vliegen, CJ Stam, SA Rombouts, and RW Keunen. "Rejection of the 'filtered noise' hypothesis to explain the variability of transcranial Doppler signals: a comparison of original TCD data with Gaussian-scaled phase randomized surrogate data sets." *Neurological research*, **18**(1):19–24, 1996.

[VSV07] SVN Vishwanathan, Alexander J Smola, and René Vidal. "Binet–Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes." *International Journal of Computer Vision*, **73**(1):95–119, 2007.

[VV07] Michel Verhaegen and Vincent Verdult. *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[VVB07] Michel Verhaegen, Vincent Verdult, and Niek Bergboer. "Filtering and System Identification: An Introduction to using MATLAB Software." *Delft University of Technology*, **68**, 2007.

[WBA12] Bo Wahlberg, Stephen Boyd, Mariette Annergren, and Yang Wang. "An ADMM algorithm for a class of total variation regularized estimation problems." *IFAC Proceedings Volumes*, **45**(16):83–88, 2012.

[WL07] Matthew Ward and Jeremy A Langton. "Blood pressure measurement." *Continuing Education in Anaesthesia, Critical Care & Pain*, **7**(4):122–126, 2007.

[WRK93] Jeffrey A Winfield, Philip Rosenthal, Robert K Kanter, and George Casella. "Duration of intracranial pressure monitoring does not predict daily risk of infectious complications." *Neurosurgery*, **33**(3):424–431, 1993.

[WYZ15] Yu Wang, Wotao Yin, and Jinshan Zeng. "Global convergence of ADMM in nonconvex nonsmooth optimization." *arXiv preprint arXiv:1511.06324*, 2015.

[XCS10] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. "Robust PCA via outlier pursuit." In *Advances in Neural Information Processing Systems*, pp. 2496–2504, 2010.

[XKB10] Peng Xu, Magdalena Kasprowicz, Marvin Bergsneider, and Xiao Hu. "Improved noninvasive intracranial pressure assessment with nonlinear kernel regression." *Information Technology in Biomedicine, IEEE Transactions on*, **14**(4):971–978, 2010.

[YL06] Ming Yuan and Yi Lin. "Model selection and estimation in regression with grouped variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(1):49–67, 2006.

[YLY11] Lei Yuan, Jun Liu, and Jieping Ye. "Efficient methods for overlapping group lasso." In *Advances in Neural Information Processing Systems*, pp. 352–360, 2011.

[ZDP13] Jun Zhong, Manuel Dujovny, Hun K Park, Eimir Perez, Alfred R Perlin, and Fernando G Diaz. "Advances in ICP monitoring techniques." *Neurological research*, 2013.

[ZH05] Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(2):301–320, 2005.

[ZH18] Guoqiang Zhang and Richard Heusdens. "Distributed optimization using the primal-dual method of multipliers." *IEEE Transactions on Signal and Information Processing over Networks*, **4**(1):173–187, 2018.

[ZLS12] Caoxie Zhang, Honglak Lee, and Kang Shin. "Efficient distributed linear classification algorithms via the alternating direction method of multipliers." In *Artificial Intelligence and Statistics*, pp. 1398–1406, 2012.