Approximate credibility intervals on electromyographic decomposition algorithms within a Bayesian framework

Olivier Thill

A thesis submitted for the degree of Doctor of Philosophy

School of Computer Science and Electrical Engineering

University of Essex

July 2021

(This version was generated: October 2021)

Abstract

This thesis develops a framework to uncover the probability of correctness of algorithmic results. Specifically, this thesis is not concerned with the correctness of these algorithms, but with the uncertainty of their results arising from existing uncertainty in their inputs. This is achieved using a Bayesian approach. This framework is then demonstrated using independent component analysis with electromyographic data.

Blind source separation (BSS) algorithms, such as independent component analysis (ICA), are often used to solve the inverse problem arising when, for example, attempting to retrieve the activation patterns of motor units (MUs) from electromyographic (EMG) data.

BSS, or similar algorithms, return a result but do not generally provide any indication on the quality of that result or certainty one can have in it being the actual original pattern and not one strongly altered by the noise/errors in the input.

This thesis uses Bayesian inference to extend ICA both to incorporate prior physiological information, thus making it in effect a semi-blind source separation (SBSS) algorithm, and to quantify the uncertainties around the values of the sources as estimated by ICA. To this end, this thesis also presents a way to put a prior on a mixing matrix given a physiological model as well as a re-parametrisation of orthogonal matrices which is helpful in pre-empting floating point errors when incorporating this prior of the mixing matrix into an algorithm which estimates the un-mixing matrix.

In experiments done using EMG data, it is found that the addition of the prior is of benefit when the input is very noisy or very short in terms of samples or both. The experiments also show that the information about the certainty can be used as a heuristic for feature extraction or general quality control provided an appropriate baseline has been determined.

Table of Contents

Notation

The following notation conventions will be used throughout this thesis. An example is given each time after the colon.

- Matrices will be bold uppercase letters: $\mathbf{M}$
- Vectors will be bold lowercase letters: $\mathbf{v}$
- Constant scalars will be uppercase letters: K
- Other scalars will be lowercase letters: k
- Double barred letters denote domains of values: $\mathbb{Q}$
- Slanted letters denote models or particular functions: $\mathcal{L}$
- A circumflex denotes an estimate of the variable: $\hat{\mathbf{W}}$
- A tilde denotes a distribution a variable is drawn from: $\tilde{\mathbf{Y}}$
- A caron denotes a variant of a variable such as a test set: $\check{\mathbf{X}}$

As exceptions, $p(...)$ refers to a probability density function and Π and π refer to a Bayesian posterior or prior respectively.

Indexing of matrices and vectors will be in square braces after the variable. Indices will be given in column major order and ranges are indicated with a colon between the beginning and end which are omitted if the range is from the beginning or to the end. For example: $\mathbf{X}[k{:}K, 1]$

1 Introduction

Any real world measurement contains errors. Any model which is a simplification of a real world process, will, on average, produce results less accurate than its inputs by virtue of its nature as a simplification. Thus, the results from algorithm processing such measurements and or making use of simplifying models are, despite being mathematically sound, not necessarily the 'truth'. They may be, but one cannot be certain about this.

Aims

This thesis aims to develop a framework to uncover the probability of correctness of algorithmic results. Specifically, this thesis is not concerned with the correctness of these algorithms, but with the uncertainty of their results arising from existing uncertainty in their inputs.

There is a large class of algorithms which, given some input value, output some other value linked through some relation to the input. If the input is free from any errors, such as the ones, for example, arising when recording real world data, and the model employed to relate between input and output is completely accurate, then the output can be expected to be free from error with a probability of one.

If, however, the input or the model employed or both contain inaccuracies, then the probability that the output is correct must be lower than one. That is because a mere algorithm, and not an oracle, is used to compute it.

This thesis focuses on the probability that the output of such an algorithm is the actual "truth" given an unknown amount of uncertainty introduced by the input date and the models employed inside the algorithm.

This class of algorithms often returns a singular result. In the case of algorithms that employ some iterative method, the statement still holds true when treating the threshold at which they stop as a limitation of their capability to model the relation between input and output. That singular result is generally, within the domain of possible results, the most probable one given the data and the internal model employed the algorithm.

Often that result is all that is needed to proceed. More often than not, using it as is will not cause issues, either because the systems consuming these results are sufficiently tolerant to inaccuracies in their input, or because the data feed to the algorithm was good enough not to cause the result to deviate too far from the "truth".

However, if the latter statement does not hold, then the errors caused may be hard to find as the algorithm returns exactly that what is expected of it given its current input and model, even when that output is not the desired "truth". One way to detect this form of issue is to consider the actual probability that the result is the correct one.

Of course, the probability of specific result is, in itself, not the full story as it does not capture the probabilities of all other possible results. For that one needs to know the probability distribution over all results which is often not tractable to compute but there are a number of ways to approximate such distributions.

Even if such distribution is, approximately, known, it can be hard to interpret it at a glance. If one, however, reframes this problem in Bayesian terms, then that distribution, assuming a uniform prior distribution, would then be the Bayesian posterior distribution, then the Bayesian credible interval becomes available as a tool which can give strong insights to the question's answer.

Independent component analysis (ICA) is an example of a class of algorithms commonly used to get the most probable result without regard to how probable it truly is. This thesis aims to demonstrate how ICA, as an example, can be extended to provide information about the Bayesian credible interval around its result in addition to the result itself.

For the purpose of this example, ICA will be used to solve the inverse problem encountered when working with electromyographic (EMG) data. The latter measures the electric signals underlying the activation of muscles. It is used in neural interfaces to, among other things, control prostheses. It is also employed in medical applications to diagnose certain conditions although, in this case, the source separation step central to this thesis is not normally employed.

As noted before, it is necessary to define a prior distribution when employing Bayesian methods. If all that is desired is to unlock the ability to perform analysis with the tools of Bayesian statistics, the prior may be an uninformative one. Alternatively, it may be informative and, thus, potentially improve the algorithms results. Thus, the second objective of this thesis is to define an informative prior for the specific problem stated above and to further extend ICA to make use of it. The first objective being the development of the components of the aforementioned framework to quantify the certainty of algorithmic results given non-perfect input.

Challenges

The primary challenges are those arising from selecting an appropriate model to incorporate prior physiological knowledge and to select the model parameters and those arising from numerical instabilities appearing in direct implementations of the model and algorithms.

There is a plethora of existing models for electromyography, as discussed in section 2.2, ranging from simple to complex, computationally light to heavy and for different geometries. However, the trade-offs between ease of use and realism of results are not necessarily readily apparent for these models.

Likewise, there have been multiple studies into the electric properties of tissue. However the exact methodology and thus the results often differ. Even when the methodology is the

same, more often than not, the results, while qualitative similar, are not similar from a raw numeric perspective.

Furthermore, mathematical instabilities due to floating point error arose many times in the process of the practical work. The main issues of this kind arose when calculating a gradient, either when the forward pass involved a matrix inverse, particularly when whitening was not used, or when some of the parameters where discreet integers. The solution to the first is presented in section 3.5 and the latter was 'solved' by avoiding the need for discreet integer parameters altogether.

Thesis Structure

In this 1st chapter, the aims of this thesis and its challenges are introduced. Then the structure of the chapters is presented.

Subsequently, a thorough literature review of the relevant topics is given in chapter 2. Any relevant specific literature and prior work used in this thesis will also be addressed in this chapter. Neural interfaces, particularly those generally classified as brain computer interfaces are covered in section 2.1 with the subtopic of electromyography being additionally covered in-depth in section 2.2. Section 2.3 presents an overview of the inverse problem which is often encountered when working with electromyographic data and brain computer interfaces in general. Finally, section 2.4 covers Bayesian inference and algorithms to estimate a Bayesian posterior, all of which is heavily used in the experimental section.

Chapter 3 presents the theoretical work undertaken for this thesis and the algorithms developed. The chapter begins with section 3.1 in which the aims and challenges will be restated in more technical language. Sections 3.2 and 3.3 quickly cover how the pre-existing algorithms are used and implemented in the context of the novel algorithms presented later. Section 3.4 describes a method to derive a Bayesian prior from a physiological model in a way which is mostly independent from the actual model used. Then an algorithm to re-parametrise orthogonal matrices using Householder reflections is presented in section 3.5. This re-parametrisation allows certain optimisations and avoids mathematical instability in the algorithms from sections 3.3 and 3.6. The core work of this thesis is then presented in sections 3.6 and 3.7 where the algorithms introduced from the sections before are combined into an algorithm which allows to formulate a independent component analysis solution as a full Bayesian posterior. Furthermore, Section 3.7 also presents ways to leverage the additional information gleaned to make statements both about the quality of the algorithm output as a whole and about the quality of specific parts of the output.

The practical work is then described in sections 4.1 and 4.2 of chapter 4 in the form of a number of experiments run using the algorithms and models described in chapter 3. Then section 4.3 presents and discusses the results of the experiments run with synthetic data. This is followed by section 4.4 which does likewise for real data while contrasting these to the results of the previous section.

Finally chapter 5 concludes this thesis by summarising the findings from chapter 4 and of the thesis as a whole. Then, it enumerates possible directions of further research.

2 Literature Review

As stated in chapter 1, this thesis aims to formulate a general framework to quantify the uncertainty of results of algorithms given imperfect input and or employing imperfect models. The intent is to create this framework using Bayesian Inference, covered in section 2.4. The algorithm chosen as demonstrative example is Independent Component Analysis (ICA) which is used to solve the so-called inverse problem. Both off which are covered in section 2.3. The data used, in said demonstration, is electromyographic (EMG) data whose nature and specific challenges are covered in section 2.2. EMG is one of the technologies used in neural interfaces which are covered in section 2.1 to give a broader context of possible uses for the work presented in this thesis.

2.1 Neural Interfaces

Aims and Applications

Neural Interfaces record physiological signals originating, directly or indirectly, from the neural systems and translate them into input for another system. There exist a wide variety of sensor types for the capture of the brain signals as well as a wide selection of algorithms to process them depending on what is intended to be achieved.

In terms of aims, neural interfaces can be roughly categorised into a number of categories. Systems which aim to restore some physical ability that a person lost through some event, e.g accident, or that a person was born without [1], but that are normally found in a healthy human [2]. Systems which aim to enhance human capabilities beyond their natural potential [3]. Systems aimed for entertainment purposes [4]. Systems whose purpose is to monitor a persons brain or similar activity either to aid in neural and or psychological research [5] or to diagnose certain conditions [6] as well as to help with certain rehabilitation procedures [7].

Neural interfaces which involve computer processing are generally described as brain computer interfaces (BCI) even when the interface is with a neural system outside of the brain.

Restore Some of the most common systems aimed at restoring lost functionality are the ones that allow to control prostheses [8, 9, 10]. In the ideal case a user of a brain-controlled prosthesis would have as much control over it as over a natural body part. However, technical limitations currently restrict the degrees of freedom that such prostheses provide to around ten or less, but devices allowing for significantly more degrees of freedom are in the research stage [11]. Related research also looks into giving sensory feedback to the user of these

prostheses [12, 13]. There has also been some pioneering work to restore limited vision to people who became blind after birth [14].

Another type of brain computer interface often encountered in the literature are spelling systems in which the user selects a series of symbols, e.g. letters or numbers, through certain mental processes and thus can spell out messages [15] or control a computer mouse pointer [16]. Such interfaces can be one of the few means that fully locked-in patients can use to communicate with the outside world [17]. However, some studies on this were criticised as it is quasi impossible to validate the correctness of the "communications" in some such cases [18].

Enhance While not as popular of an aim, as judged by the number of publications, most of the research to restore lost functionality also applies to enhancing those capabilities in healthy humans. For example exoskeletons can help persons with muscular affliction but also help to increase a person's strength beyond human standards [19].

Research is also been conducted to enhance various mental capacities [20]. An interesting trial to enhance the correctness of decision making of a group is presented in [21] where, among other things, features concerning the users certainty in their own performance were used.

Entertain In the last years, a number of games that are meant to be played via input from neural interfaces have been developed. The first ones were mostly based on the same approaches as the aforementioned spellers or motor-imagery [4]. Since then, most techniques generally used for neural interfaces have been used for such games [22].

Additionally, there exist games which do not use the neural interface for direct control, but to affect the game state based on the users mental and emotional state [23].

Research Neural interfaces are used extensively as a research tool. Research into stuttering [24], tinnitus [25], effects of childhood trauma [26], effects of meditation [27] and the general workings of the brain [28] represent only some samples of such research.

Diagnosis One of the earliest uses of equipment to observe neural signals was to diagnose various conditions and illnesses such as epileptic seizures [29] or chronic pain [30] and it is still used for these purposes [31, 32].

Current research is often focused on augmenting or even replacing the analysis of the recorded signals with various machine learning methods [33, 34] or on discovering currently unknown patterns and markers indicating some condition [35, 36].

Rehabilitation In addition to restoring functionality through neural interfaces, it has been found that the use of such systems while giving visual [7] or haptic [37] feedback helps

to retrain and rebuild damaged neural systems even if a patient, such as a stroke patient, isn't currently able to physically move their muscles.

Further to this, the same information, as gleaned from neural interfaces for diagnosis, remains also of use during any rehabilitation process [38].

General Workflow

The general outline of most BCI systems can be described as a sequence of steps. Note that only the first and the last step, as described below and shown in figure 2.1, are found in all BCI applications; all other steps are optional and may be omitted in a given implementation. The separation into discrete steps here is for exposition only; in a practical implementation, most of these steps, if implemented at all, may well be combined or ordered differently. Furthermore the exact methods employed during each step are often dependent upon whether the system is online or offline. Offline here means that the data was fully recorded before processing starts while online means that the system is processing the physiological signals in real time as it is recorded.

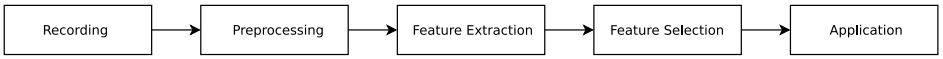

Figure 2.1: The general outline of most BCI systems can be described as a sequence of steps.

Recording In the first step, the data is recorded using appropriate equipment. Digitalisation and some basic filtering may also be applied here. E.g. a band stop filter to mitigate the interference of the mains supply [39].

Preprocessing The second step consists in the application of one or more advanced preprocessing passes whose aim is to increase the signal to noise ratio, which tends to be low for BCI data, as much as possible by removing latent noise and by detecting areas of the data which have been overly corrupted by so-called artefacts [40].

Feature Extraction The third step attempts to reduce the dimensionality of the input by condensing the information carried by the data into so-called features. This is often achieved by extracting statistical properties of the data or transforming the domain of the data, using for example the Fourier transform [41].

In this step, if running an online algorithm, windowing may be also be applied to the data, i.e only the last n samples are selected and passed on to the next steps [42].

Feature Selection If the third step does not reduce the data sufficiently, then it is possible to compute the average usefulness of each feature for the subsequent steps in a fourth step and drop all but the most useful [43].

Applications In the fifth step the processed data is given to a predictor, which may be a classifier or a regressor, of some description which attempts to estimate the neural state of the subject [44].

Finally, something is done with the data. What exactly depends, of course, on the purpose of each BCI system.

Recording Technologies

There are several different technologies to capture the raw signals as produced by the brain or the efferent nervous system [45]. These technologies can be categorised according to a variety of features. A major distinction can be made between those technologies considered invasive, i.e. they require some form of surgery, and those that do not. Other features distinguishing BCI technologies are the spatial and temporal resolutions of the recordings as well as the portability and general user-friendliness of the recording process [46]. Finally, a distinction exist between those technologies that record brain activity directly, those that record it indirectly and those that record the peripheral nervous system directly or indirectly as opposed to the central one.

Technology to affect the brain through applied stimulus can also be considered as divided between invasive and surface applications.

Surface Recording Technologies

One of the most commonly measured phenomenon is the electric potential or, more precisely, its change over time as the average magnitude is subject to a multitude of unknown factors such as the nature of the electrodes, their contact with the skin and the concrete composition of the tissue below them. This varying electric potential is produced by ion gradients propagating along the nerve cells in the case of electroencephalography (EEG) [47] or along muscle tissue as is the case in electromyography (EMG) [48].

All cells produce an electric potential between there insides and their outside by selectively allowing ions to through their surrounding membranes. Some cells, primarily neurons and muscle fibres, can change the ion concentrations around their membranes rapidly, this is called depolarisation, when subject to an electric stimulus such as produced by such a depolarisation event happening in close proximity. For details on this mechanism see [49, Chapter 21].

EEG is probably the most well known technology due to its, relative to other technologies, low price, safety and ease of use without a specialised lab [50]. It is still preferable to record

in a shielded laboratory on an immobile patient but it is possible to gather more noisy but still useful data without those constraints as opposed to other technologies. Moreover, the shielding needed for recordings of electric potential can be provided by a simple Faraday cage [51].

Recordings are done by placing electrodes in a grid pattern on the scalp, often with the help of a cap. The grid positions for EEG have been standardised in the so called 20-10 system though extensions for larger numbers of electrodes do exist [52]. Each electrode is expected to record the sum of all activity in the brain, mostly of the cortex, in a large radius around it. As a consequence, EEG exhibits poor spatial resolution as the recoded signals are essentially a mixture of many separate underlying sources. Individual neuron activity is too faint for EEG to even register in a surface electroencephalogram: entire groups of neurons need to activate "together" in a short time-frame.

One challenge when recording EEG signals is that the skull which has a low conductivity compared to the surrounding tissue is of a somewhat complex non-homogeneous shape; this distorts the recorded signal. Taking measures of the conductive properties of the skull and scalp into account has, however, been demonstrated to alleviate this issue to some degree [53].

A similar technology to EEG is electromyography (EMG), here the electrodes are placed on the skin above the muscle and record the potentials arising from the ion gradients travelling along muscles fibres which are triggered by electric signals originating from motor neurons whose synapses end at the muscle. These gradients are what causes the muscles to contract. Like EEG, the observed EMG signals are distorted, as discussed further in section 2.2, by the tissue located between the signal origins and the recording electrodes [54].

There are no properly standardised locations for the placement of electrodes for EMG, but the number and arrangement of the sensor may also significantly affect the nature of the recording [55]. As a consequence, different EMG datasets often need different algorithms to process them. EMG will be covered in more depth in section 2.2. An example grid configuration can be found in figure 4.1.

Both EEG and EMG recordings rely on a reference electrode. There are multiple options for this: either there is a dedicated electrode whose signal is subtracted from all other electrode signals, commonly at the top of the head or the earlobes for EEG, or adjacent electrodes are subtracted in series from each other, this is called single differential, or a spatial filter such as a Laplacian is applied, or the mean of all electrodes is used as a synthetic reference electrode [56].

Magnetoencephalography (MEG) records the magnetic fields which arise from the currents measured by EEG. The main benefit of MEG over EEG is a greater spatial accuracy and, thus, resolution. Additionally, MEG only measures currents in the tangential direction relative to the sensors and it does not record signals arising from as deep in the brain as EEG, both of these features could be considered a drawback or a benefit. On the other side, MEG requires very large equipment thus is not portable. In addition, the materials needed

to build and operate MEG equipment are much more expensive than those needed by EEG or EMG [57].

Functional near infrared spectroscopy (fNIRS) and functional magnetic resonance imaging (fMRI) are both technologies used to measure the brain state indirectly by measuring the flow of blood within the brain. These methods have very good temporal and spatial resolution. However, it has to be taken into account that the phenomenon measured is only indirectly related to "thinking" activity and, thus, there is often a significant lag between the recorded signal and the underlying neural activity [58].

Invasive Recording Technologies

The main benefit of using invasive probes, despite the obvious danger to the patients' health during the installation of the sensors, is that they capture the signals close to their sources as opposed to non-invasive technologies which record the signals with sensors far from the origin of the signals. As a result invasive technologies provide the best spatial and temporal resolution among all technologies. In particular, they allow localised recordings while non-invasive methods only record large areas at once, but, reversely, invasive technologies cannot provide a recording of a large area unless very many recording elements are used which, evidently, multiplies the health risks [59].

Ethical Considerations

The use of BCI technology raises a number of ethical considerations and related issues [60].

Informed Consent Concerns

Some of its medicinal use cases are those where the ability to give consent, let alone informed consent, are made difficult by the ailment to be treated, i.e. patients who are fully locked-in cannot possibly give assent to BCI without the use of BCI, that is, no other technologies which allow them to give consent are known of at this point; which is a chicken and egg problem.

A related issue of consent arises when the condition to be treated is correlated with cognitive impairment or if the patients are too young to be considered able to give consent.

Furthermore, particularly for invasive BCI installations, the risk associated with the implantation are not well researched due to lack of samples and, in some cases, lack of consequent long-term studies of the effects of implantation where such was done.

The latter problem leads to another issue for invasive BCI research procedures. Namely, the long term cost of the implantation, as pointed out in [61]. Who pays for the maintenance of the BCI system after the study is done and the grant associated with it is used up?

Finally, there is the question on whose shoulders the responsibility of the actions of BCI controlled devices or applications lies? The user? The software? Or its authors [62]?

Privacy and Security Concerns

BCI systems are computer systems and, as such, are subjected to all the security concerns that entails. They can potentially be suborned by malicious software or transmission streams carrying brain state information could be intercepted. Potential attacks could further include an attacker taking over control of a prosthesis to do things with it, which ties into the agency and responsibility issue raised above [63].

Or an attacker could be after private information, and, while BCI by no means can read thoughts at this point, it still can reveal a number of sensitive data ranging from emotional state to low entropy information, such as which numbers the individual concentrates on while typing in a pin code [64].

2.2 Electromyography

Overview

While most brain computer interface technologies focus on gleaning information from the brain, electromyography is more geared towards observing the somatic nervous system. Of course, the state of the peripheral somatic system itself is heavily influenced by the central nervous systems' state. Electromyography records electric potentials within muscles. These potentials are a direct consequence of signals from the brain arriving via motor neurons. The benefit of this indirect recording is that the muscles act as natural amplifiers and thus provide a much stronger signal than the motor neurons themselves.

A thorough review of this topic beyond what is covered in this chapter can be found in [65].

Origin of the Signal

Electromyography records the electric fields generated during muscle contractions. Each motor neuron is connected, at so called end-plates, to a few or, depending on type, many muscle fibres. These connections are generally located close to the centre of the muscle fibre. The contractions are set off by electric impulses arriving at the end-plates which set off a cascade of ion gradients which propagate in both directions and which cause the contraction of the part of the muscle fibres over which they pass as well as the field potentials recorded by electromyography [66].

The recorded values from surface EMG originate thus from many individual points in space within the body. Each such point or cluster of closely related points tends to emit a signal distinctive, by some measure, of that point. Such points in space will be referred to as sources for the remainder of this thesis.

Application

it is possible to reconstruct, at least partially, the intended movement of amputated extremities based on electromyographic information recorded from remaining body parts, e.g. hand movement based on EMG data from the upper arm. Thus EMG is commonly used to allow control of prostheses [8].

Beyond that, a number of afflictions, such as the aftermath of strokes, can leave a patient unable to properly move their muscles while still generating sufficient EMG signal to derive their intent, which can be used to control exoskeletons which can be used for rehabilitation [10]. Of course EMG driven exoskeletons could also be used by healthy persons as a form of physical augmentation [19]. EMG data is also used as part of the diagnosis process of a variety of muscular diseases [33, 30].

Challenges

The recorded values from surface EMG originate from many individual sources inside the body thus the recorded signals are each a mixture of the signals from many physiological sources. Furthermore, the number of observations is usually significantly lower than the number of the sources [67].

Complicating the situation further, there exist numerous factors which distort the signal between source and sensors. Moreover, there are usually a significant amount of external noise factors which often cause the observed signal to have a signal to noise ratio lower than one [68].

Additionally, as noted before, the signal amplitude there is attenuated due to the properties of the tissue between the signal origin and the recording equipment [54]. If that tissue is inhomogeneous, distortion can be expected to occur. It is to note that homogeneity is assumed for simplicity in the majority of the literature.

In EMG, the attenuation was considered to be mostly governed by the resistive electrical property of the tissue, however, this is contested in more recent sources. In particular, the claim is made that the effects of the properties of capacitance and conductivity are not negligible, as previously thought, thus they must be taken into account [69]. Furthermore, the observed attenuation is not invariant to the frequency of the recorded signal [70].

As EMG is what controls the movement of muscles, in most scenarios where the EMG signal is of interest, the movement and its effects on the recorded signal are to be taken for granted. This is concerning because the movement of tissues causes the sources of the EMG signal to also move relative to the recording electrodes which may cause distortions in the recording. An exception to the above arises if the muscles have been surgically affixed to the bone, which is part of some amputation techniques [71].

As noted above, what is recorded during EMG is a mixture of signals, not the signal itself which is the actual information of interest. Thus, one has the problem of how to recover

that information. Problems where the original signal is unknown, but only the result of a transformation, whose parameters are unknown, is known, are commonly considered a variation of the inverse problem. The general inverse problem and the algorithms used to recover the source signal from the observation are covered in section 2.3.

Tissue Properties

The physiological properties of muscle and other body tissues of primary concern to EMG are those affecting the travel of the electrical potentials along the muscle fibres and how the electric fields created from these potentials extend to outside the skin where they are recorded.

Considerable research has been done on this properties within the field of bioelectromagnetics which is primarily concerned with the interactions between electromagnetic fields and biological matter. One of its main areas of research being dosimetry which is concerned with how electric and magnetic fields affect living biological tissue. To make assumptions on said effects without in-vivo experiments, it is necessary to know the electrical properties of the tissues of interest as this allows to run experiments in simulations [72].

There are three main tissues, not counting blood, through which an EMG signal will travel between source and electrode. These are:

- Muscle, which has a high conductivity and thus a high amplitude loss. Note that, in simplified terms, conductivity is a measure of how much free electrons can move around and the energy used for that movement comes "out" of the amplitude. Furthermore, it is anisotropic [73] which means that an electric field changes amplitude differently depending on the direction of the transversal [74]. Various measurement studies have been done, such as [75, 76], but often come to slightly different results as can be seen in the comparisons found in [77].

- Fat, which has lower conductance than muscle, thus lower loss. As it is above the muscle layer and thus above all sources, anisotropy is much less an issue as the direction of the traversal is much more predictable [74].

- Skin, whose various layers have very varied properties which makes it hard to determine their effects [78].

Muscle has a high conductivity and thus a high amplitude loss. Note that, in simplified terms, conductivity is a measure of how much free electrons can move around and the energy used for that movement comes "out" of the amplitude. Furthermore, it is anisotropic [73] which means that an electric field changes amplitude differently depending on the direction of the transversal [74]. Various measurement studies have been done, such as [75, 76], but often come to slightly different results as can be seen in the comparisons found in [77].

Fat has lower conductance than muscle, thus lower loss. As it is above the muscle layer and thus above all sources, anisotropy is much less an issue as the direction of the traversal is much more predictable [74].

The various layers of the skin have very varied properties which makes it hard to determine their effects [78]. In the experiments presented in chapter 4 the values of these properties come from [79] and, as there, the capacitive effects are ignored to allow for a better comparison of results in section 4.1.

Further discussion in detail of these properties as pertaining to EMG can also be found in [65].

Motor Units

A motor unit (MU) is the group of nerve-endings which originate from one motor neuron together with the muscle fibres enervated by these [80]. Ion gradients cascade from the sites of innervation in one MU along the muscle fibre when triggered by an incoming impulse from the MU's motor neuron. These moving gradients give rise to measurable electric potentials, each called a motor unit action potential (MUAP) [81]. In the remainder of this thesis, what will be meant when talking about an EMG signal is the waveforms of such MUAPs when recorded over time, unless otherwise specified.

These MUAPs propagate primarily along the length of a muscle fibre. This happens at a speed of approximate around $4\frac{m}{s}$ to $8\frac{m}{s}$, although outliers do exist. The exact values depend heavily on the composition of the muscle which in turn is a function of a given subjects lifestyle [82, 83]. If sampling EMG data at around 16kHz, this results in an approximate move of the source between 0.25mm per sample to 0.5mm per sample.

Furthermore, MUAPs are not point processes. Their amplitude over time follows a distinct general shape akin to a hat as shown in figure 2.2. The duration of this waveform roughly measured—as where to start and end is not standardised—was around 15ms in one particular experiment [84]. This would put the duration of one such MUAP at around 240 samples long at a sampling rate of 16kHz. Note that the waveform of one MUAP is also commonly refereed to as a spike.

Multiplying the approximate speed of propagation and the approximate duration results in the assumption that the tail of a MUAP is around 9cm away from its head; this is subject to the accuracy of the data derived from the literature as presented in the previous paragraphs. Nevertheless, it can be assumed thus that a given MUAP can be recorded across a large area at any given point in time albeit during different phases of the spike shape.

Recording Hardware

If the recording electrode or electrodes are inserted into the muscle itself, one talks of invasive EMG. As touched upon before, the benefits of invasive EMG is the high signal to noise ratio

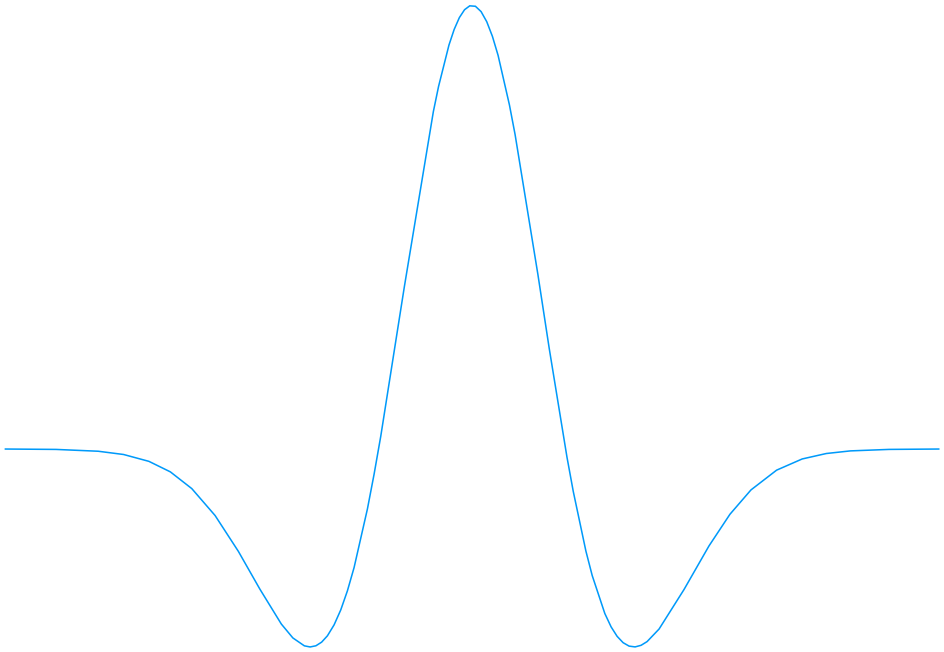

Figure 2.2: Approximate MUAP shape over time.

of the signal and the low distortion of the signal due to the small distance between the sources and the electrode. The disadvantage is that invasive EMG, while less dangerous than other invasive BCI technologies, still causes minor damage to the body [85].

Given the above considerations, invasive EMG is useful to investigate a single motor unit MU at a time as each motor unit's MUAPs can be separated with relative ease from other MU's [86]. However, for the same reason, invasive EMG can only give information about a small region and, thus, unless a great many electrodes are used simultaneously—which would multiply the health risks—cannot be used to gain a general image of the activity across a whole muscle.

The non-invasive form of EMG recording is known as surface EMG. If only one electrode is used then what is recorded is essentially the average activity of all MUs in the vicinity of the electrode and it is impossible to separate the individual MUAPs [87].

Given the limitations of single channel surface EMG, nowadays, EMG is often recorded using arrays of electrodes rather then a single one. This is referred to as HD-EMG. The additional and partly redundant information of such a setup helps to mitigate some of the drawbacks due to it being a non-invasive recording technique. In particular, it is possible to attempt to separate distinct MUAPs from each other [88, 89]. This will be addressed further in section 2.3.

Mixture Models

Numerous models for EMG can be found in the literature. Their main differences are, firstly, whether they are derived analytically or based on finite element models and, secondly, the

closeness of their models to reality. The latter being a function of the geometry of their models and of how many distinct tissue layers they simulate.

While these models can be used to determine the electric potentials at any point in the volume conductor, the information sought in the context of this thesis is how a potential at a source manifests at the location of a recording electrode as this mapping determines the mixing matrix described in section 2.3.

Assuming that the contributions of different sources are cumulative but independent, then models $\mathcal{M}$, such as presented later in this section, can map the relative locations of sources and electrodes to a mixing matrix $\mathbf{M}$ given parameters Θ. These parameters are derived according to the electric properties of a volume conductor which is parametrised to approximate the muscle, fat and skin tissue located between the presumed locations of the sources and the electrodes.

$$\mathbf{M} = \mathcal{M}(\Theta) \tag{2.1}$$

Each element of the resulting mixing matrix $\mathbf{M}$ represents the contribution that one source makes to the value of one electrode. The row i of $\mathbf{M}$ is composed by the contributions of each source to the value of the ith electrode. The column j of $\mathbf{M}$ represents the total of contributions to the values of all electrodes made by the jth source. More precisely, the element of $\mathbf{M}$ at row i and column j represents the contribution of jth source to the value of the ith electrode. The rationale behind this will be covered in section 2.3.

Analytical Modelling

Analytical models are those that can be expressed in closed form; concretely, that means that given an analytical model $\mathcal{M}$ and some input parameters Θ—which here include the values of the sources and the locations of both sources and electrodes—then the value of $\mathcal{M}(\Theta)$ can be calculated per hand or with a computer without having to resort to approximations or iterative algorithms such as gradient descent based machine learning.

The most basic approach would be to assume that each source would give contribution equal to its value to each electrode:

$$\mathbf{M}[i, j] = 1 \tag{2.2}$$

Evidently, while a valid mixing matrix could be derived from such a model, this would be very far from the truth as it can be assumed with certainty that the contributions of a given source to different electrodes differ according to some factor:

$$\mathbf{M}[i, j] = 1 \cdot \mathbf{A}[i, j] \tag{2.3}$$

where $\mathbf{A}[i, j]$ is the coefficient describing both how the contributions of the j source to the ith electrode differ compared to its contributions to other electrodes and how the contribution received by ith electrode from the jth source differ from the contributions it receives from all other sources.

Assuming a purely homogeneous volume conductor, the contributions of any source to any electrode should depend purely on the distance between their locations and a fixed attenuation factor α [90]:

$$\mathbf{M}_{i,j} = 1 \cdot e^{-\alpha||\upsilon_j - \chi_i||} \tag{2.4}$$

where υ_j and χ_i are the locations of the ith source and jth electrode respectively, and α is a constant derived from the electric properties of the volume conductor:

$$\alpha = \omega\sqrt{\frac{\mu'\epsilon'}{2}\Big(\sqrt{1 + \big(\frac{\sigma_{\text{eff}}}{\omega\epsilon'}\big)^2} - 1\Big)} \tag{2.5}$$

where μ' and ϵ' are the real parts of the magnetic permeability and electric permittivity respectively, and where ω is the assumed angular frequency of the electric current, and where σ_{eff} is the effective conductivity. Furthermore, it is to note that, in the absence of any magnetic field of significance, which is the normal case, $\mu' = \mu_0$, where μ_0 is the know constant of vacuum permeability [72].

However, the model described above still indicates a significantly lower amount of attenuation compared to the prediction from very detailed models such as [55]. Taking into account the border conditions between different tissue types improves the results from the model compared to what is expected. The equations to do so can be found in [72].

A range of analytical models can be found in the literature. These range from very simple models assuming an infinite volume of muscle [91] to models with cylindrical geometry comprised of multiple layers for each distinct tissue. A one-layer model can be found in [92], a two-layer extension of that in [93] and finally a three-layer version is described in [94]. A variant of this using a three-layer planar models can be found in [95, 96]. These models all present a different trade-off between precision and complexity. [95] is used in the practical part of this thesis due to being not overly complex to implement while still remaining relatively close to reality.

Finite Element Modelling

Analytical models which try to be very close to reality can become very complex. Thus, the approach of Finite Element Modelling (FEM) is used in those cases. In the context of modelling EMG, FEM splits the entire volume conductor into many small connected cells with simple geometries. Then it is attempted to find by iteration a solution to the equations given above, or similar ones depending on the setup, for each such cell. The equations are much simpler for each cell as opposed to the whole volume conductor as each cell can be mostly treated as homogenous. However, the large number of cells makes FEM still a very computationally expensive method and thus not suited for the work underdone in this thesis. A general introduction to FEM can be found in [97] and two examples of FEM models for EMG can be found in [98] and [55].

2.3 The Inverse Problem

Problem Description

The problem where multiple sources are observed in a mixed manner at multiple points is the base problem dealt with by blind source separation (BSS) algorithms. Assuming that the mixture process is linear, then the problem may be described with the following equations:

$$\mathbf{X} = \mathbf{MY} \tag{2.6}$$

$$\mathbf{Y} = \mathbf{WX} \tag{2.7}$$

where $\mathbf{Y}$ is a matrix containing the sources and $\mathbf{X}$ is a matrix containing the observations where the number of columns of $\mathbf{X}$ and $\mathbf{Y}$ is the number of samples in the data. Note that the equations remain valid even with only one sample where $\mathbf{X}$ and $\mathbf{Y}$ would be column vectors.

$\mathbf{M}$ is the matrix describing how those sources are mixed and $\mathbf{W}$ is its inverse, that is $\mathbf{M}^{-1}$. Note that the above equations don't take into account the noise inherent to any real recordings and imply that neither $\mathbf{W}$ nor $\mathbf{M}$ are singular, i.e. they are invertible.

The aim of BSS algorithms is to estimate $\mathbf{M}$ directly or indirectly by estimating its inverse $\mathbf{W}$. A multitude of algorithms have been developed in the last decades. They differ by the assumptions on the properties, such as independence, Gaussianity, sparsity or non-negativity, of the data they attempt to exploit [99]. Those algorithms that employ situation specific information beyond the heuristics listed above are generally classified as semi-blind source separation (SBSS).

Spatial Convolution A descriptive example for the inverse problem is a room full of talking people and a series of microphones. What is recorded at each microphone at each moment is a mixture of all voices at that moment. With voices from people further from that microphone being more faint than those originating from a closer location. In other words, it is a convolution over the spatial domain centred on the microphone given some convolution matrix with the inputs being the voices of the people.

If one would have the recordings of all microphones and wanted a transcript of what each person said, then one needs to reverse that convolution. To do so, is to solve the inverse problem.

The above example can be easily transferred into the domain of BCI. For EEG and EMG for instance, one needs just to replace the microphones with the respective electrodes and the people with MUs or neuron clusters respectively, with the signal to recover being their respective action potentials [100]. The latter is what is attempted in the experiments described in chapter 4.

Temporal Convolution The equations 2.6 and 2.7 imply that the mixture model is instantaneous and, thus, the temporal domain is of no consequence. However, this is, in all real world cases, a simplification of reality because information can never travel faster than light [101]. Thus, signals emitted at one point in time from two different sources don't arrive at the same time at the point of observation unless the sources are exactly equal in distance from the point of observation.

Assuming all sources emit a signal at the same time. Then let the maximum lag L be the difference in time between the first detection of any of these signals at any electrode and the last detection of any of these signals at any electrode. Then the mixture process of sources $\mathbf{Y}$ to observations $\mathbf{X}$ could be described by:

$$\mathbf{X}_t = \sum_{l=0}^{L} \mathbf{M}_l \mathbf{Y}_{t-l} \tag{2.8}$$

If L would be 0 because the maximum difference is much less than the sampling period, then the above equation simplifies to those of an instantaneous mixture.

However, this representation is not very practical. In [102] a model is proposed where lagged versions of the observations, both past and future, are "added" to the observations. The benefit of not only considering how the sources relate to the observed mixture at one given moment in time, but also how the sources affect the mixture at points in the past and future is to essentially increase the number of observations by treating past and future observations as separate observations. As the number of sources which, mathematically, can be identified is proportional to the number of observations, having more observations allows more sources to be identified from a given mixture.

A different approach is taken in [103] where, iteratively, the observations in $\mathbf{X}$ are shifted based on correlation between channels recoded at adjacent electrodes, then one source is recovered and, subsequently, removed from the observations before starting the next iteration.

Another aspect concerning the temporal dimension is that time series often have specific temporal structures, such as the one shown in figure 2.2, which can be exploited to guide the separation process by SBSS algorithms [99, ch.19].

Over and under-determined cases

If there are more sources than observations then the problem is under-determined. In this case it is not mathematically possible to recover all sources and the sources recovered by an algorithm might still be mixtures of multiple sources [104].

If, on the other hand, there are more observations than sources, then the problem is over-determined. Generally this only means that more information is available than mathematical necessary which is a "good thing". However, some algorithms—e.g. symmetric ICA, see section 3.2—don't deal well with this if the number of sources is unknown [104].

In practise, such as with the EMG data used in chapter 4, it can happen that the problem is technically under-determined but the data is so noisy that only an amount of sources less

than the number of observations can be recovered. Such scenario was found by the author to behave like the over-determined case for all practical purposes.

Spike Detection

One form of BSS algorithms for EMG observations are those that attempt to extract MUAP traces by seeking evidence for the characteristic "spike" pattern of MUAPs.

Template Matching

One way to separate the sources from the observed mixture is to find as many as possible templates of what the signal looks like based on spikes which were not strongly distorted by the mixture process. Then one iterates over the data and when a match is found, it is recoded and the matched template is subtracted from the mixture. This is repeated until no other matches are found. The remaining signal is being assumed to be part of the background noise [105]. Note that this algorithm could be considered to be a SBSS algorithm as it makes use of knowledge about the general shape of such spikes.

Principal and Independent Component Analysis

The following algorithms are related mathematically and conceptually. Principal component analysis (PCA) could be understood as a precursor to independent component analysis (ICA) although it has use cases outside of BSS too. Whitening is a required preprocessing step for ICA which can, among other ways, be derived from PCA.

Principal Component Analysis

In PCA, a dataset is transferred to a new orthogonal basis such that a maximum of variance information is contained in the first n features of the transformed dataset where n can be any number smaller or equal to the number of features in the dataset. This is achieved through decorrelation, i.e. making the covariance matrix zero in all non diagonal elements [106].

For the PCA algorithm to work, the assumption that any noise is i.i.d and more Gaussian than the data of interest must hold true and, ideally, the noise should have a covariance matrix of identity [106].

Let $\mathbf{X}$ be a centred data matrix with n rows and m columns:

$$\mathbf{\Sigma}_{\vec{e}}\mathbf{\Sigma}_e\mathbf{\Sigma}'_{\vec{e}} = \mathbf{\Sigma} = \frac{\mathbf{X}\mathbf{X}'}{m-1} \tag{2.9}$$

where $\mathbf{\Sigma}$ is the covariance matrix of $\mathbf{X}$ and $\mathbf{\Sigma}_e$ is the diagonal matrix containing the eigenvalues of $\mathbf{\Sigma}$ ordered by descending order of magnitude and $\mathbf{\Sigma}_{\vec{e}}$ the corresponding eigenvectors stacked column-wise:

$$\mathbf{X}_{\text{PCA}} = \mathbf{\Sigma}'_{\vec{e}}\mathbf{X} \tag{2.10}$$

Whitening

The whitening transform, sphering in some literature, is a combination of decorrelation and variance rescaling such that the covariance matrix of the data becomes the identity matrix.

There are infinite potential whitening matrices for any given dataset as any arbitrary rotation or reflection of whitened data results in whitened data. Thus, the whitening matrix is only defined up to an multiplicative orthogonal matrix.

There are five common algorithms to derive a whitening matrix, each with their own effects on $\mathbf{X}$ when applied to it [107].

Let $\mathbf{\Sigma}$, $\mathbf{\Sigma}_{\vec{e}}$ and $\mathbf{\Sigma}_e$ be as defined above, let $\mathbf{C}$ be the corresponding correlation matrix, $\mathbf{V}$ be the diagonal variance matrix, and $\mathbf{C}_{\vec{e}}$ and $\mathbf{C}_e$ be to $\mathbf{C}$ as $\mathbf{\Sigma}_{\vec{e}}$ and $\mathbf{\Sigma}_e$ are to $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = V^{\frac{1}{2}}\mathbf{C}\mathbf{V}^{\frac{1}{2}} \tag{2.11}$$

Given that notation the equations of these algorithms are:

- $\mathbf{W}_{\text{pca}}$ is a direct extension of PCA and retains its utility in regards to ease of dimensionality reduction:
$$\mathbf{W}_{\text{pca}} = \mathbf{\Sigma}_e^{-\frac{1}{2}}\mathbf{\Sigma}'_{\vec{e}} \tag{2.12}$$
- $\mathbf{W}_{\text{zca}}$ (Zero Phase Component) minimises the cross-covariance between the whitened and original data:
$$\mathbf{W}_{\text{zca}} = \mathbf{\Sigma}_{\vec{e}}\mathbf{W}_{\text{pca}} = \mathbf{\Sigma}^{-\frac{1}{2}} \tag{2.13}$$
- $\mathbf{W}_{\text{pca\_cor}}$ is as $\mathbf{W}_{\text{pca}}$ except in regards to $\mathbf{C}$:
$$\mathbf{W}_{\text{pca\_cor}} = \mathbf{C}_e^{-\frac{1}{2}}\mathbf{C}'_{\vec{e}}\mathbf{V}^{-\frac{1}{2}} \tag{2.14}$$
- $\mathbf{W}_{\text{zca\_cor}}$ is as $\mathbf{W}_{\text{zca}}$ except in regards to $\mathbf{C}$:
$$\mathbf{W}_{\text{zca\_cor}} = \mathbf{C}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}} \tag{2.15}$$
- $\mathbf{W}_{\text{chol}}$ does not optimise any particular property of the whitened data but tends to be in-between any of the previous options of whitening matrices in regards to any of the specified properties:
$$\mathbf{L}\mathbf{L}' = \mathbf{\Sigma}^{-1} \tag{2.16}$$
$$\mathbf{W}_{\text{chol}} = \mathbf{L}' \tag{2.17}$$
where $\mathbf{L}\mathbf{L}'$ is the result of a Cholesky decomposition. Note that
$$\check{\mathbf{L}}\check{\mathbf{L}}' = \mathbf{\Sigma} \tag{2.18}$$
$$\mathbf{M}_{\text{chol}} = \check{\mathbf{L}}' \tag{2.19}$$
where $\mathbf{M}_{\text{chol}}$ is the inverse of $\mathbf{W}_{\text{chol}}$. This permits to use this form of the decomposition without explicitly calculating the inverse of $\mathbf{\Sigma}$.

Independent Component Analysis

One of the most common BSS algorithms is ICA which uses the central limit theorem to separate sources by attempting to minimise the Gaussianity of the sources because a mixture is always more Gaussian than its sources [108].

However, there are other qualities instead of Gaussianity which can be used as measure if it can be assumed that the sources exhibit this distinct quality strongly [99].

One such quality is sparsity. If it is known that the signals from the sources are very sparse, that is zero most of the time, then, this property can be used instead of entropy as a measure to be maximised inside the ICA setup [109].

Another possible measure is the non-negativity of the original factors, that is the sources are assumed to be non-negative which can be said to often model real world data well [110].

Pre-Whitening Most ICA algorithms converge only slowly, if at all, if the given input is not centred and white, i.e. the mean of each row is 0 and $\mathbf{XX'} = \mathbf{I}$. To sidestep this issue, $\mathbf{X}$ is generally first centred and then premultiplied by a whitening matrix $\mathbf{W}_0$ before applying ICA. The result of ICA then is a matrix $\mathbf{W}_1$ from which the actual un-mixing matrix $\mathbf{W}$ can be recovered through the simple relation: $\mathbf{W} = \mathbf{W}_1\mathbf{W}_0$. The centring could also be undone after un-mixing but this is neither commonly done nor necessarily a sensible thing to do.

The sources estimated by ICA are always uncorrelated in addition to having their independence, by some measure, maximised [111]. Thus:

$$\mathbf{W}_1\mathbf{W}_0\mathbf{X}\mathbf{X}'\mathbf{W}_0'\mathbf{W}_1' = \mathbf{I} \tag{2.20}$$

and since by the definition of the whitening transform

$$\mathbf{W}_0\mathbf{X}\mathbf{X}'\mathbf{W}_0' = \mathbf{I} \tag{2.21}$$

$\mathbf{W}_1$ must be an orthogonal matrix as

$$\mathbf{W}_1\mathbf{I}\mathbf{W}_1' = \mathbf{I} \tag{2.22}$$

Maximum Likelihood ICA If $\mathbf{X}$ is a S by T matrix, then the mathematical equation for maximum likelihood ICA is [104]:

$$\mathcal{L}(\mathbf{W}|\mathbf{X}) = \prod_t^T \prod_s^S p(\mathbf{W}[s,:]\mathbf{X}[:,t])||\mathbf{W}|| \tag{2.23}$$

$$p(x) = 1 - \tanh(x)^2 \tag{2.24}$$

where t indexes the samples in time and s is an index for the rows of $\mathbf{W}$ which themselves produce the sth source in $\mathbf{Y}$. $||\mathbf{W}||$ is the absolute determinant of $\mathbf{W}$. The function p given here can be any measure of non-Gaussianity, the one given here is merely one of the most commonly used. The ICA solution of $\mathbf{W}$ is that which is the most likely, i.e. it maximises the above likelihood function.

FastICA FastICA, as the name implies is, often, one of the fastest ICA algorithms to find $\mathbf{W}_1$. There are two general categories of this algorithm: the deflationary ones which estimate one row of $\mathbf{W}_1$ after the other and the symmetric ones which estimate all rows of $\mathbf{W}_1$ simultaneously [112].

The symmetric kind estimates all rows of $\mathbf{W}_1$ simultaneously, which, theoretically, minimises the error in $\mathbf{W}_1$ due to the noise in $\mathbf{X}$ but it does not give the sources in any particular order [113] and has been observed by the author to exhibit very bad performance if the number of actual sources is less than the number of sources sought for.

On the other hand, the deflation-based ICA algorithms estimate one source after the other which prevents issues if there are less sources than rows in $\mathbf{W}_1$. However, estimation errors accumulate for each estimated source and can become significant for the last sources to be estimated [113]. If the number of actual source equals the number of sources sought for, deflation based ICA thus tends to perform worse but it is much more robust when the number of actual sources is less than the number of sources sought for as all the "error" would accumulate in the final sources which are only "noise" anyway.

As the errors in deflation-based ICA stem mostly from the errors in the already estimated sources, the order in which they are found can play a significant role in determining the quality of the complete output. However, that order is defined by the value by which $\mathbf{W}_1$ was initialised to, which, in the standard ICA algorithm, is done randomly. In reloaded ICA, the ideal order is first estimated from a $\check{\mathbf{W}}_1$ itself estimated through any appropriate algorithm, and then ICA is run with an accordingly permuted version of $\check{\mathbf{W}}_1$ as input [114, 115]. Note that this makes the algorithm more robust towards noise but still does not account for it like most ICA algorithms.

Blind VS Semi-Blind Source Separation

BSS algorithms traditionally attempt to minimise the amount of prior information needed to successfully separate mixtures. Prior information in BSS algorithms is thus usually limited to the assumptions inherent in the algorithms themselves as described above. However, there are situations where the use of more prior information is useful or even necessary for the accuracy of the decomposition. Thus, several semi-blind source separation algorithms have been developed [99, ch.19].

A well-known example of SBSS algorithms are those based on Bayesian theories which will be covered more in section 2.4. If using this approach for SBSS, there is a well-defined mathematical way to introduce prior knowledge into the equation through the appropriately named prior π. The posterior Π can be considered an improved version of our prior knowledge as it takes into account the observed values. In particular, if more observations would become available, it is possible to use the old posterior as the prior for the current iteration [99, ch.16].

Work attempting to use prior information for SBSS of BCI data has been done to some extent using combinations of EEG and MEG [109, 116, 117]. However, for EMG, while

significant work has been done to create models based on physiological information, those models are mostly presented as a means to validate other algorithms, which do not use prior information themselves but rather are being used for source separation themselves [65]. An exception to this can be found in [118], [119] and [120], however, only a limited amount of prior information was used there. From the same author is [121] where the prior information is used in classification not separation. Another paper of interest is [122] where the possible utility of subject specific information for the algorithm is discussed in the conclusion.

Independent component analysis has also been embedded into Bayesian frameworks in the past while working with EMG data. However, no physiological information was used to determine the prior nor was any in-depth analysis of the posterior or posterior predictive distributions performed [123].

2.4 Bayesian Inference

Bayes theorem [124], which is at the core of all Bayesian algorithms, describes the relationship between the conditional probabilities $p(\Theta|\Xi)$ and $p(\Xi|\Theta)$ as:

$$p(\Theta|\Xi)p(\Xi) = p(\Xi|\Theta)p(\Theta) \tag{2.25}$$

which is also often put into this form:

$$p(\Theta|\Xi) = \frac{p(\Xi|\Theta)p(\Theta)}{p(\Xi)} \tag{2.26}$$

When used in Bayesian inference, Θ is the set of parameters to be estimated and Ξ the data which carries (new) information about Θ. As a rule in Bayesian statistics, the data Ξ is considered fixed and the parameters Θ to be random variables [125].

The probability density function $p(\Theta)$ is known as the prior π, which is the distribution of Θ according to what is known before the new information in Ξ is considered. Analogously $p(\Theta|\Xi)$ is known as the posterior Π which is the distribution of Θ according to what is known after the new information in Ξ is considered. $p(\Xi|\Theta)$ represents the likelihood $\mathcal{L}$ that Θ is the truth given the Ξ. Note that $p(\Xi|\Theta) = \mathcal{L}(\Theta|\Xi)$, i.e. the arguments are "flipped". The distribution $p(\Xi)$ is often unknown and not easily computable, including through marginalisation, and thus, as it is essentially a constant since Ξ is constant, often dropped. This leaves the following general formula:

$$\Pi(\Theta|\Xi) \propto \mathcal{L}(\Theta|\Xi)\pi(\Theta) \tag{2.27}$$

The posterior Π is a probability distribution over the whole domain of Θ. If only the value of Π at the mode of this distribution, usually called the maximum a posteriori (MAP), is required then this can be formulated as an optimisation problem:

$$\mathrm{MAP}(\Theta|\Xi) = \arg\max_{\Theta} \Pi(\Theta|\Xi) \tag{2.28}$$

or

$$\arg\max_{\Theta} \Pi(\Theta|\Xi) \sim \mathcal{L}(\Theta|\Xi) \cdot \pi(\Theta) \tag{2.29}$$

The later can often be solved with "simple" gradient descent based algorithms instead of the probabilistic algorithms needed to compute an estimate of Π. Both types of algorithm will be covered below.

Gradient Descent

In many applications it becomes desirable to maximise or minimise certain properties. In mathematical notation this optimising problems are written as:

$$\arg\max_{\Theta} \mathcal{F}(\Theta) \tag{2.30}$$

for maximisation and

$$\arg\min_{\Theta} \mathcal{F}(\Theta) \tag{2.31}$$

for minimisation respectively where $\mathcal{F}$ is a scalar valued function with input arguments Θ.

If the function of interest $\mathcal{G}$ is scalar valued and the desired property does coincide with $\mathcal{G}$'s minima or maxima, then $\mathcal{G}$ can be used directly as $\mathcal{F}$, else a so-called loss function which has the desired properties can generally be interposed, i.e. the $\mathcal{F}(\Theta)$ from the above equations would actually be $\mathcal{F}(\mathcal{G}(\Theta))$.

Gradient descent algorithms operate by calculating the first derivative or partial derivatives if Θ is not scalar and calculate from this the gradient Δ, that is the direction in which Θ would need to change for $\mathcal{F}(\Theta)$ to increase. That is

$$\mathcal{F}(\Theta) < \mathcal{F}(\Theta + \eta\Delta) \tag{2.32}$$

is assumed to hold provided the learning rate η is sufficiently small. This small change of Θ done iteratively until the magnitude of Δ falls below a threshold, is, in essence, the complete core of gradient descent.

The choice of η and how to change it after each iteration is the primary difference between gradient descent algorithms and still a field of active research. An overview of current algorithms can be found in [126].

Probabilistic Algorithms

There often do not exist close form solutions for Bayesian posteriors. Various algorithms to sample from or to estimate Π have been developed despite this. They are presented below in descending order of accuracy, which is also their descending order of computational complexity.

Markov Chain Monte Carlo

MCMC, which arises out of the combination of Markov chains and Monte Carlo sampling method, is a somewhat computational expensive but otherwise conceptually simple and stable method.

It is a family of algorithms which create Markov chains whose elements distribution converges to a given stationary distribution irrespective of the initial value of the chain provided it is sufficiently large [127].

Markov Chain A Markov process is characterised by the property that the current state of the modelled process depends only on the immediately preceding state; otherwise formulated, the next state only depends on the current state. Any prior states are considered uncorrelated or represented fully through the current state already. The transition between two states is described by a so-called Markov kernel. In simple words, the kernel is a function which takes the current state as parameter and returns the next state. This kernel function may be either deterministic or probabilistic as required by the modelled process [128].

A special case are hidden Markov processes, also referred to as the latent Markov model, where a hidden state follows the Markov process, but one can't observe said state directly. What can be observed is a variable whose value depends on the value of the current hidden state according to a known mathematical relationship [129]. A simple visual representation for this is:

$$\begin{array}{ccccc} x_1 & \rightarrow & \cdots & \rightarrow & x_n \\ \Downarrow & & \Downarrow & & \Downarrow \\ y_1 & & \cdots & & y_n \end{array}$$

where x_1 through x_n represents the values of a unobserved Markov process with a kernel function $\rightarrow$ and y_i represents the observed values of x_i according to a transfer function $\Downarrow$.

Monte Carlo The Monte Carlo sampling method refers to the random sampling from one or many distributions followed by deterministic processing. Its use derives from the fact that it is often much simpler to simulate many singular instances of a complex system and to draw conclusions from that sample than it is to describe the expected output of a complex system given only a description of the system itself [130].

MCMC uses Monte Carlo sampling to determine the next state of a Markov chain such that, on average, after many samples, the chain becomes distributed like the sought posterior distribution.

Variants The classical variant of MCMC is Metropolis Hastings [131]. Here, for each step, if the current state is x, a random new state y is proposed and accepted with probability

$$p(x \rightarrow y) = \min\left(1, \frac{\Pi^*(y)}{\Pi^*(x)}\right) \tag{2.33}$$

where Π^* is the, possibly unnormalised, posterior in the case of Bayesian inference. If the move to y is not accepted then the old state x is chosen as the new state.

Hamiltonian Monte Carlo (HMC) is a modern variant of MCMC which uses gradient information in conjunction with the namesake Hamiltonian equations [132] to more efficiently sample from the full posterior domain than the random walk employed by Metropolis Hastings. Even a basic description of the concrete method employed in HMC is too large to include here; an introduction can be found in [133].

Variational Bayes

Variational Bayes (VB) is an algorithm to produce an approximation of the posterior. In VB each parameter in Θ is assigned to a parametrised probability distribution family. In a common approach, called mean field, independence between each parameter is assumed. Then the parameters of these distributions are found by some method, iteratively or simultaneously, such that the Kullback-Leibler divergence between the resulting distribution and the posterior is minimised [134].

Laplace Approximation

Laplace's method is an algorithm to evaluate integrals by approximation. Provided the maximum a posteriori has been estimated with sufficient accuracy, then the mathematics of that method can be used to find the moments of a normal distribution centred on the MAP. Those moments can be used to derive a reasonable approximation of the posterior [135].

Predictive Distributions

A predictive distribution in the Bayesian context is the distribution of unobserved values $\check{\Xi}$ given observed values Ξ where the distribution of $\check{\Xi}$ depends on Θ. As Θ is drawn from a distribution itself, the predictive distribution, to take into account Θ's distribution, is first calculated by treating Θ as variable and then subsequently Θ is marginalised out.

The posterior predictive distribution is calculated as:

$$p(\check{\Xi}|\Xi) = \int p(\check{\Xi}|\Theta, \Xi)\Pi(\Theta|\Xi)\,\mathrm{d}\Theta \tag{2.34}$$

The prior predictive distribution is calculated as:

$$p(\check{\Xi}|\Xi) = \int p(\check{\Xi}|\Theta, \Xi)\pi(\Theta)\,\mathrm{d}\Theta \tag{2.35}$$

2.5 Summary

This chapter reviewed the topic of neural interfaces and its subtopic of electromyography. Usage of EMG data often stipulates that an inverse problem is to be solved. Namely what

the underlying MUAP signals are given the mixture recorded at the electrodes. The inverse problem in general and the independent component analysis algorithm in particular have been covered in section 2.3 of this chapter. Then this chapter ended with a review of Bayesian inference, which will be used in the next chapter to provide a framework to analyse the certainty of the out put of ICA. This framework will then be applied to EMG data in chapter 4.

3 Bayesian Extension to ICA

This chapter will describe extensions made, as part of the work towards this thesis, to the independent component analysis (ICA) algorithm and procedures developed to that end. The main extensions lie in incorporating prior knowledge and providing a measure of quality using a Bayesian framework. Unless explicitly stated otherwise, it can be assumed that everything presented in this chapter is original work or refers to observations made by the author himself in the course of the practical work for this thesis.

3.1 Overview

All algorithms will be presented in the context of the inverse problem posed by electromyographic recordings. However nothing about them, except the specific physiological model employed, is strictly specific to EMG, i.e. the algorithms and methods employed here ought to be adaptable to any instance of the inverse problem with very little work.

Aims

The ultimate practical use of every algorithm, or set of algorithms, tackling the inverse problem is to recover the signal $\mathbf{Y}$ which was transformed by some mixing process $\mathbf{M}$ into the observations $\mathbf{X}$. It is assumed that this is a linear mixing process, thus this corresponds to finding that processes inverse $\mathbf{W}$.

Let $\mathbf{Y}$ be represented by a S_y by T matrix where each of the S_y rows represents a source, which in the case of EMG data is a trace of a MU activation pattern, and where T is the number of samples over time. Similarly, let $\mathbf{X}$ be represented by a S_x by T matrix where each row represents an observation, one row per electrode in the case of HDsEMG. Then the matrix $\mathbf{W}$, known as un-mixing matrix, is a S_x by S_y rectangular matrix and $\mathbf{M}$, the mixing matrix, is a S_y by S_x matrix. For mathematical convenience S_y is often given the same value as S_x, thus making $\mathbf{W}$ a square matrix which allows $\mathbf{M}$ to be a straightforward matrix inverse of $\mathbf{W}$.

In the noiseless case, these values relate to each other through these simple equations:

$$\mathbf{X} = \mathbf{MY} \tag{3.1}$$

$$\mathbf{Y} = \mathbf{WX} \tag{3.2}$$

Independent Component Analysis aims to find an estimate $\hat{\mathbf{W}}$ such that each source in the reconstructed $\hat{\mathbf{Y}}$ takes on certain properties as discussed in section 2.3. It however neither reports on its own accuracy nor can it be guided by any information except those aforementioned domain-independent properties.

The goals of the here presented extension are to make the overall algorithm produce a measure on its accuracy which does not depend on the user manually inspecting $\hat{\mathbf{Y}}$ and to improve the results of the algorithm by injecting prior knowledge, based on anatomy and electrophysiology, about $\mathbf{M}$ and thus $\mathbf{W}$.

While these two goals are independent to each other per se, they can be achieved simultaneously by embedding ICA, specifically its maximum likelihood formulation, into a Bayesian framework. Thus they will be covered together here.

Challenges

There are two main challenges.

Firstly all ICA algorithms suffer from an ambiguity of scale and order of the sources estimated through them. The scale ambiguity is mitigated in the maximum likelihood formulation through normalisation such that the standard deviation of all estimated sources is 1. The permutation ambiguity can be resolved by defining some ordering criteria. Such reordering is however not amenable for integration into gradient descent methods.

Secondly it is reasonably simple to define a prior on $\mathbf{M}$ but much harder to define it directly on its inverse $\mathbf{W}$. All attempts, made pursuant this thesis, to define the prior directly on $\mathbf{W}$ were unsuccessful, though further research might find a way. As a consequence, in the following, the prior is defined on $\mathbf{M}$ which implies an inverse of the parameter to be optimised which can lead to mathematical instabilities. This can be addressed by noting that $\mathbf{W}$ may be split into $\mathbf{W}_1$ and $\mathbf{W}_0$, where $\mathbf{W}_0$ is constant given $\mathbf{X}$ and $\mathbf{W}_1$ mathematically ought to be an orthogonal matrix whose inverse thus simply is its transpose. However, while some ICA algorithms such as FastICA ensure the orthogonality of $\mathbf{W}_1$ after each update step by renormalising it, the here used maximum likelihood formulation does not usually enforce this. Thus, to use iterative algorithms, such as gradient descent or Markov Chain based sampling algorithms, the $\mathbf{W}_1$ part of $\mathbf{W}$ needs to be re-parametrised such that the orthogonal quality of $\mathbf{W}_1$ is preserved after each step.

Methods

To begin, estimates $\hat{\mathbf{W}}$ and $\hat{\mathbf{Y}}$ are computed from $\mathbf{X}$ using reloaded FastICA [115] as described in section 3.2. Presuming that all formulations of ICA give the same result, assuming the same measure of independence is used, then $\hat{\mathbf{W}}$ should be equivalent, up to permutation, to the maximum likelihood solution of ICA and under the assumption of a flat prior even to the maximum a posteriori solution. Let the latter concept be refereed as $\hat{\mathbf{W}}_{\mathcal{L}}$ in the following.

Then a gradient descent algorithm is used as described in section 3.6 to find the maximum a posteriori (MAP) solution $\hat{\mathbf{W}}_{\Pi}$ with an informative prior as described in section 3.4 and a maximum likelihood function as described in section 3.3. This is sped up significantly by

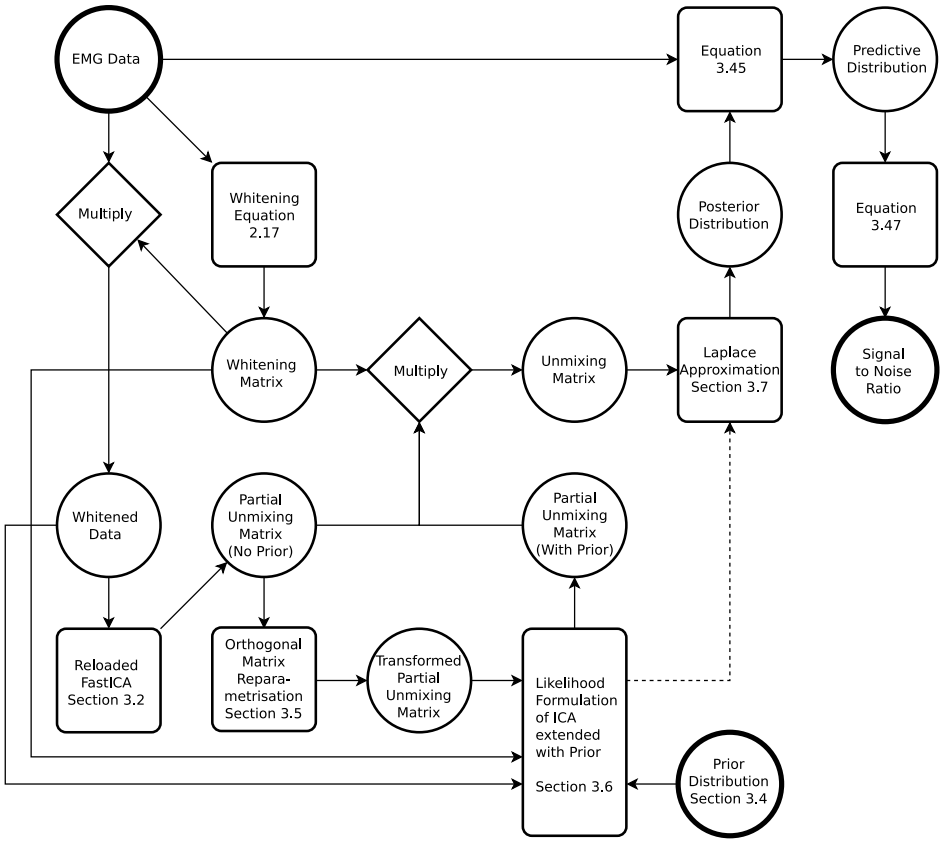

Figure 3.1: Overview of the framework presented in chapter 3. Circles denote variables and squares denote computations or algorithms. Note that some implementation details have been omitted for overall clarity.

using $\hat{\mathbf{W}}_{\mathcal{L}}$ as a starting point under the assumption that it is very close to one of the global maxima.

The full posterior distributions for both $\hat{\mathbf{W}}_{\mathcal{L}}$ and $\hat{\mathbf{W}}_{\Pi}$ is then estimated as described in section 3.7. With estimates of these posterior distribution, the posterior and maximum likelihood predictive distributions of $\mathbf{Y}$ are estimated. As the estimated predictive distributions are still subject to the scaling ambiguity of ICA, a signal to noise ratio of each source can be calculated as a last step as also described in section 3.7.

3.2 Finding $\mathbf{W}_1$

Data of high definition electromyography recorded from S electrodes, not considering grounds, consists of S time-series of T samples each. Let such data be $\mathbf{X}$ and let $\mathbf{Y}$ be the unmixed sources, which are approximately the MUAP trains. Both can be represented as matrices of S rows and T columns. This assumes a number of unmixed sources equal to the number of observations. In practice, a number of sources in $\mathbf{Y}$ will be false sources containing only or mostly noise.

$\mathbf{X}$ and $\mathbf{Y}$ are assumed to be linear mixtures of each other. Assuming that noise can be

ignored, then it follows from the equation of the inverse problem in section 2.3 that if the un-mixing matrix $\mathbf{W}$ or its inverse the mixing matrix $\mathbf{M}$ is known, then, $\mathbf{Y}$ can be fully recovered from $\mathbf{X}$. See also equations 3.1 and 3.2.

$\mathbf{W}$ can be decomposed into two parts:

$$\mathbf{W} = \mathbf{W}_1\mathbf{W}_0 \tag{3.3}$$

similarly for $\mathbf{M}$:

$$\mathbf{M} = \mathbf{M}_0\mathbf{M}_1 \tag{3.4}$$

where $\mathbf{M}_0$ and $\mathbf{M}_1$ are also the inverses of $\mathbf{W}_0$ and $\mathbf{W}_1$ respectively. This only holds if these matrices are square, which is the primary reason the number of unmixed sources is taken to be equal to the number of observations.

The spilt of $\mathbf{W}$ is usually performed because most algorithms have significantly better convergence characteristics when searching for $\mathbf{W}_1$ given a pre-whitened $\mathbf{X}_0$ calculated as:

$$\mathbf{X}_0 = \mathbf{W}_0\mathbf{X} \tag{3.5}$$

This split will be, however, of further use in section 3.6.

Let $\mathbf{W}_0$ be a whitening matrix of $\mathbf{X}$. That is a matrix such that

$$\operatorname{cov}(\mathbf{W}_0\mathbf{X}) = \mathbf{I} \tag{3.6}$$

then there are infinitely many possible $\mathbf{W}_0$ as

$$\operatorname{cov}(\mathbf{Q}\mathbf{W}_0\mathbf{X}) = \mathbf{I} \tag{3.7}$$

holds true for all $\mathbf{Q} \in \mathbb{Q}$. Multiple algorithms to derive $\mathbf{W}_0$ from $\mathbf{X}$ are presented in section 2.3. As the choice of $\mathbf{W}_0$ is not particularly important for independent component analysis the Cholesky method is chosen here for its simplicity and relative numerical stability.

As discussed in section 2.3, all versions of independent component analysis return sources which are uncorrelated, thus the covariance matrix of the estimated $\mathbf{Y}$ is diagonal and the covariance matrix is the identity matrix in effect because the variances are usually normalised to 1 as ICA cannot recover scale.

It follows thus that $\mathbf{W}$ itself is expected to be a whitening matrix and $\mathbf{W}_1$ thus must be an orthogonal matrix if $\mathbf{W}_0$ is a whitening matrix.

With FastICA

If no prior information is to be taken into account, then the FastICA family of algorithms, discussed in section 2.3, is one of the fastest ways to estimate $\hat{\mathbf{W}}_1$ given $\mathbf{W}_0$. Convergence when using $\mathbf{W}$ and $\mathbf{X}$ directly however is very slow, and has been observed by the author to fail outright.

Symmetric FastICA returns the sources in an arbitrary order subject to the random initialisation of $\hat{\mathbf{W}}_1$. And the quality of the estimate $\hat{\mathbf{Y}}$ tends to fare badly in the presence of noise if there are less actually recoverable sources than the mathematical maximum as implied by the rank of $\mathbf{W}$, i.e. the number of rows in $\mathbf{X}$. Since the number of recoverable sources in $\mathbf{X}$ can not be expected to be known a priory but is expected to be regularly less than the rank of $\mathbf{W}$, this is troublesome, because it makes the symmetric approach not suitable for situation described here.

The basic deflatory FastICA algorithm also does not order sources in a deterministic manner like the symmetric algorithm. However, it is robust against cases where there are less recoverable sources than rows in $\mathbf{X}$ as it recovers the sources sequentially.

Reloaded FastICA, as described in [115], is an extension to deflatory FastICA which explicitly orders the sources according to a source dependent metric. This does not guarantee the same ordering when for example the algorithm is applied to different parts, or the same part preprocessed differently, of a dataset, but, as long as it can be assumed that the same sources are recovered, it can be assumed that the ordering is at least very similar. This is the algorithm which was employed in the experiments described in chapter 4.

Source Matching

As stated, Independent Component Analysis returns the found sources in an arbitrary order which is generally not predictable or comparable between different inputs and sources may have their sign flipped, i.e. are multiplied element-wise by minus one. More formally, ICA only finds, approximately, $\mathbf{P}\mathbf{M}'_1$ where $\mathbf{P}$ is any signed permutation matrix.

If source estimates of the same data have been found through different ICA algorithms or with different parameters, or a ground truth is know because synthetic data was used, then one of the estimates needs to be reordered to match the other, that is the differences between the unknown $\mathbf{P}$s introduced by the ICA algorithm need to be found.

To do this an heuristic must be chosen which signals how likely two estimated sources represent the same source. Good but simple candidates for this are the absolute covariance or the absolute correlation. The absolute value is needed as the sources may be flipped. Taking the square of these would also be a possibility.

The values of that heuristic can be combined into a matrix $\mathbf{H}$. To match the sources, a permutation matrix $\check{\mathbf{P}}$ needs to be found such that the trace of $\check{\mathbf{P}}\mathbf{H}$ is maximised. In the case where a ground truth is matched, this $\check{\mathbf{P}}$ is the transpose of $\mathbf{P}$ thus cancelling each other fully out, as permutation matrices are orthogonal matrices. This can be treated as an instance of the stable marriage problem [136] for which the Gale–Shapley algorithm [137] is an ideal solution in the sense that it returns an ordering where it is not possible to improve the result for one source through further swapping without decreasing the result for another.

Once $\check{\mathbf{P}}$ has been found, possible sign inversions of the matched sources can be corrected by flipping the sign of the appropriate element of $\check{\mathbf{P}}$ where the correlation between two

matched sources is negative.

As a note, this $\check{\mathbf{P}}$ can be directly applied to an estimate of $\mathbf{W}$ without changing the validity or quality of that $\mathbf{W}$ as it only changes the ordering and the sign of the sources, both of which ICA is invariant to.

3.3 Maximum Likelihood ICA

It is possible to formulate ICA as a maximum likelihood problem. The naive likelihood equation for a S by T dataset $\mathbf{X}$ as presented in section 2.3 is

$$\mathcal{L}(\mathbf{W}|\mathbf{X}) = \prod_{t}^{T}\prod_{s}^{S}(1-\tanh(\mathbf{W}[s,:]\mathbf{X}[:,t])^2)||\mathbf{W}|| \tag{3.8}$$

The maximum likelihood estimate $\hat{\mathbf{W}}_{\mathcal{L}}$ could then be found by utilising a gradient descent optimiser such as Nesterov-accelerated adaptive moment estimation (NADAM) [138].

Performance of the above in a gradient descent algorithm would however be poor. To have a performant implementation $\mathbf{X}$ needs to be pre-whitened and it is generally faster to compute everything in the log domain. As noted before, $\mathbf{W}_0$ is $\mathbf{W}_{\text{chol}}$ as presented in section 2.3.

To avoid awkwardness of notation, the following rule will be observed in the equations below: On the left hand side the mathematical parameters for which the function is defined are used. On the right hand side the decomposed or composed forms as would be passed in an actual implementation are used. Mainly $\mathbf{W}$ is $\mathbf{W}_1\mathbf{W}_0$ and $\mathbf{X}_0$ is $\mathbf{W}_0\mathbf{X}$.

With whitening the equation becomes

$$\mathcal{L}(\mathbf{W}|\mathbf{X}) = \prod_{t}^{T}\prod_{s}^{S}(1-\tanh(\mathbf{W}_1[s,:]\mathbf{W}_0\mathbf{X}[:,t])^2)\cdot||\mathbf{W}_1||\cdot||\mathbf{W}_0|| \tag{3.9}$$

which in the log domain is

$$\log\mathcal{L}(\mathbf{W}|\mathbf{X}) = \sum_{t}^{T}\left(\sum_{s}^{S}\log(1-\tanh(\mathbf{W}_1[s,:]\mathbf{W}_0\mathbf{X}[:,t])^2)\right)+\log||\mathbf{W}_1||+\log||\mathbf{W}_0|| \tag{3.10}$$

considering only the part within the inner sum for now with the input to tanh considered to be one variable x

$$\log(1-\tanh(x)^2) \tag{3.11}$$

then expand tanh

$$\log\left(1-\frac{(e^x-e^{-x})^2}{(e^x+e^{-x})^2}\right) \tag{3.12}$$

lift the 1 into the numerator and expand all terms in the numerator

$$\log\left(\frac{e^{2x}+2e^xe^{-x}+e^{-2x}-e^{2x}+2e^xe^{-x}-e^{-2x}}{(e^x+e^{-x})^2}\right) \tag{3.13}$$

cancel out terms

$$\log\left(\frac{2e^{x}e^{-x}+2e^{x}e^{-x}}{(e^{x}+e^{-x})^{2}}\right) \tag{3.14}$$

$e^{x}e^{-x}$ is always one

$$\log\left(\frac{4}{(e^{x}+e^{-x})^{2}}\right) \tag{3.15}$$

propagate the log

$$\log(4)-2\log(e^{x}+e^{-x}) \tag{3.16}$$

$e^{x}+e^{-x}$ is equal to $2\cosh(x)$

$$\log(4)-2\log(2\cosh(x)) \tag{3.17}$$

expand and cancel terms a last time

$$-2\log(\cosh(x)) \tag{3.18}$$

and emplace into the log-likelihood

$$\log\mathcal{L}(\mathbf{W}|\mathbf{X})=T(\log||\mathbf{W}_1||+\log||\mathbf{W}_0||)-2\sum_{t}^{T}\sum_{s}^{S}\log\cosh(\mathbf{W}_1\mathbf{X}_0[s,t]) \tag{3.19}$$

When implementing the above, both $\log||\mathbf{W}_0||$ and $\mathbf{X}_0$ can be precomputed using $\mathbf{M}_0$ instead of $\mathbf{W}_0$ to avoid the inversion of $\mathbf{M}_0$ needed to find $\mathbf{W}_0$, which is the last step of the algorithm employed to find $\mathbf{W}_0$. Similarly $\mathbf{W}_1\mathbf{X}_0$ would be computed once and the whole term inside the sum vectorised. Additionally $\log||\mathbf{W}_1||$ is always zero provided the orthogonality constraint on $\mathbf{W}_1$ is maintained as described in section 3.6.

3.4 Prior Model

In this section it is discussed how to derive a prior distribution of the mixing matrix $\mathbf{M}$ given some physiological model $\mathcal{M}$.

Prior Information

It is possible, a priory, to make assumptions about a number of physiological properties as discussed in section 2.2. Primarily these are: the electric properties of the materials of the volume conductor which are the muscle including its anisotropy, the subcutaneous fat layer and the skin; the thickness of the latter two; and of the locations of MUAPs within the muscle.

For the purposes of the generation of a prior discussed here, only the relative position of a source, which is approximately a MUAP, to a recording electrode will be of relevance. Let x denote the direction parallel to the muscle fibres, y the direction orthogonal to x while remaining parallel to the surface and z the depth which is then orthogonal to both x and y.

For simplicity of parametrisation, z shall be the depth inside the actual muscle only, with the total depth being $z_t = z + S + F$ where S and F are the thickness of the skin and the subcutaneous fat layer respectively. These values are illustrated in figure 3.2.

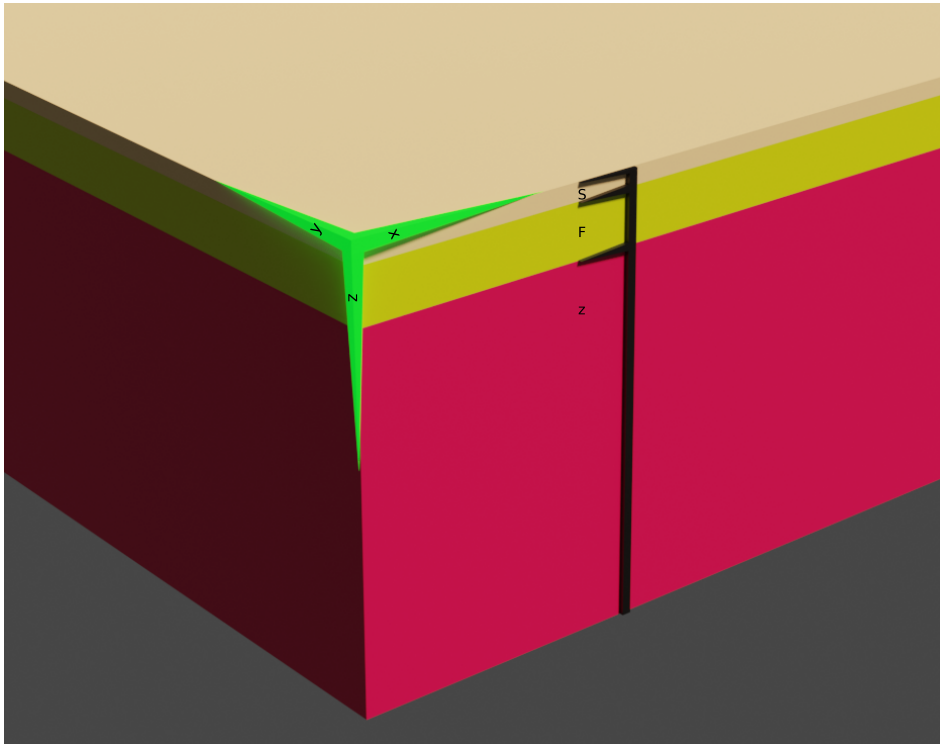

Figure 3.2: A diagram of a planar volume conductor with a skin, fat and muscle layer. On the left the axes are overlaid and on the right the lengths along the z axis are illustrated.

The conductivities for skin, subcutaneous fat and muscle shall be denoted as σ_S, σ_F and σ_M respectively. As the conductance of muscle differs along the fibre direction respective to across the fibre direction, a ratio of anisotropy A is also defined.

The following values, taken from [79], will be used for these parameters:

- $\sigma_S = 4.55 \times 10^{-4} \frac{S}{m}$
- $\sigma_F = 3.79 \times 10^{-2} \frac{S}{m}$
- $\sigma_M = 2.46 \times 10^{-1} \frac{S}{m}$
- $A = 5$
- $S = 5 \times 10^{-4} m$
- $F = 2.5 \times 10^{-3} m$

These are fixed values instead of distributions, as would more accurately describe our state of knowledge, because the computational cost would increase significantly in that case.

As discussed in section 2.2, effects of the tissue's capacitance may be, and is here, assumed to be negligible. Also as discussed in that section, these values are reported with highly variable values. Thus the concrete values above are quoted here for reproducibility purposes only.

Prior Distribution Generation

Each column of $\mathbf{M}$ describes the contribution of one source to the observations at all electrodes. One can thus, as electric fields are not assumed to interfere with each other, assume that each column of $\mathbf{M}$ is i.i.d. provided that the locations of all sources can be assumed to be also distributed i.i.d. according to some distribution $\tilde{S}$. This assumption makes the prior distribution of $\mathbf{M}$ fully definable by a prior distribution of its columns.

Given a volume conductor with a skin thickness of S and a fat layer thickness of F and the electric properties as defined above, let $\mathcal{M}$ be a model which predicts the amplitude β recorded at an electrode if there is an active motor unit action potential located at an offset $(x, y, S + F + z)$ of it with a signal amplitude α.

Then, given an electrode grid configuration, a column of a mixing matrix can be estimated for any given MUAP location within the muscle. Essentially the mixing matrix for one source located at $(x, y, S + F + z) \sim \tilde{S}$ is a column vector $\mathbf{m}$ with values derived as

$$\mathbf{m}[i] = \mathcal{M}(|x - x_{\chi_i}|, |y - y_{\chi_i}|, S + F + z) \tag{3.20}$$

where x_{χ_i} and y_{χ_i} are the grid coordinates of the ith electrode along the same axes as used by the model.

Using the above equation with many sample locations allows to calculate a sufficient amount of samples of columns of $\mathbf{M}$ that a distribution for such a column can be fitted over them. And as discussed above, the prior distribution of a mixing matrix of rank n is merely the combination of n i.i.d. columns whose values are such distributed.

Implementation Considerations

The mathematics of modelling the dynamics of the electric fields involved become much more tractable in the frequency domain as opposed to the spatial domain. As a consequence many models, including the model in [95] used for this thesis, involve a Fourier transform over the x–y plane for any given depth z.

It would consume significant computational power to recompute this for each sample MUAP location. This can be avoided by pre-computing a number of layers and then interpolating between them for each given sample MUAP location. Due to radial symmetries, only the volume between $0+S+F$ and $Z+S+F$ on the z axis, 0 and Y on the y axis, 0 and X on the x axis, needs to be precomputed, where Y and X are the furthest a MUAP location is assumed to be from an electrode on the x and y axes respectively. This optimisation can

be made as negative offsets in the x and y axis result in the same values as if they would be positive due to the aforementioned symmetries.

The model defined in [95] has a singularity at $x = 0, y = 0$ before the Fourier transform step. It is possible to attempt to impute it but even small errors can cause negative values after the Fourier transform, which is not possible as it is safe to assume that the lowest possible amount of signal, emitted from a source located at infinity, would be 0.

To resolve this, it was chosen to simply set the problematic value to 0 and to address the missing 0 Hz signal at the end by normalising the values. For this normalisation it is assumed that the furthest point $(X, Y, S + F + Z)$ gives no detectable signal any more, that is, it has a value of 0, and that the closest point $(0, 0, S + F)$ would have a perceived signal strength of 1. This is achieved by an element wise subtraction of the value at $(X, Y, S + F + Z)$ followed by an element-wise division of the value then found at $(0, 0, S + F)$.

Furthermore, the precomputed model represents a discrete three dimensional grid in its contained volume, not the full volume. Provided the step size between grid points is small enough, this can be overcome by applying an interpolation algorithm. Cubic interpolation was found to cause artefacts, in particular overshooting values if the sample source was located near an electrode in the x and y axes, while no such issues where detected with plain trilinear interpolation. The latter was thus used in the experiments described in chapter 4.

3.5 Orthogonal Matrix Reparametrisation

Orthogonal matrices have a number of particular properties. For a full description see [139]. The important properties in the context of this thesis are: that they are square matrices composed of orthonormal (i.e. orthogonal and with unit length) rows and columns, and that their inverse is also their transpose.

The property of orthogonality is expected of some parameters in certain algorithms. For example, the $\mathbf{W}_1$ returned from ICA algorithms is always expected to be orthogonal if the data was whitened with some $\mathbf{W}_0$. In FastICA, this is enforced by normalising $\mathbf{W}_1$ after each step. However, it is difficult to enforce other constraints on $\mathbf{W}_1$ for, among others, this reason.

In the extended ICA algorithm described below in section 3.6, a prior probability is put on $\mathbf{M}$, that is $\mathbf{M}_0\mathbf{M}_1$. As $\mathbf{M}_0$ is constant for a given dataset, it is very desirable for $\mathbf{W}_1$ to remain orthogonal at each step beyond the desire to stay within the original constraints of ICA because it makes the needed inversion of $\mathbf{W}_1$ into $\mathbf{M}_1$ significantly less troublesome in terms of numerical stability and computational speed.

It is not possible for probabilistic software to directly sample from the space of orthogonal matrices nor could gradient descent algorithms preserve this constraint on the parameters to be updated. Thus, if such frameworks or algorithms are employed while a orthogonally constraint must be satisfied, then the variable must be re-parametrised such that any parameter

value in the new space corresponds to a valid orthogonal matrix. Such a reparametrisation from an orthogonal matrix $\mathbf{Q}$ to a safe to update triangular matrix $\mathbf{T}$ is described here.

Orthogonal matrices of rank n are a representation of the orthogonal group, which is a symmetry group, see [140] for a primer on such groups. One of the properties of symmetry groups is that the multiplication of any two of its members returns another member of that group, possibly one of the members multiplied if the other one is the identity, i.e. an identity matrix. Thus there exists an infinite amount of possible decompositions for every orthogonal matrix.

The salient point is that there are always a number of ways to decompose an orthogonal matrix into simpler ones, which can be themselves described in a more compact way. For example, each orthogonal matrix of rank n can be decomposed into at most $\frac{n}{n-1}$ Givens rotation matrices plus one matrix determining the determinant. Each Givens matrix can be represented by a scalar value between 0 and 2π radians and an axis implied by its position in the decomposition.

The angular values described in the approach given above are amenable to gradient descent provided the value is interpreted modulo 2π. Nevertheless, the approach is not of use when the distribution of the parameter is desired as that distribution could be split awkwardly around the discontinuity around 2π which would cause a number of mathematical and convergence issues. This could be alleviated by manually rotating the range of the angle, but this would likely cause its own issues.

Another way to decompose an orthogonal matrix is to decompose it into Householder matrices which themselves can be represented as vectors of a length equal to the rank of the matrix.

The space of these vectors also contains a discontinuity at all vectors which are representable as $\alpha \cdot \mathbf{e}_1$ where $\alpha \in \mathbb{R}^+$ and $\mathbf{e}_1$ is a unit vector parallel to the first axis of a Cartesian coordinate system. However, this discontinuity does not risk splitting the mass of a probability thus it can be mostly ignored for most intends and purposes. Furthermore, the space contains a singularity at the zero vector which is also negligible as it can be arranged to have a probability of zero.

Derivation

Let $\mathbb{Q}^K$ be the domain of all orthogonal matrices of rank K and let $\mathbb{T}^K$ be the domain of sets $\{\mathbf{t}_k\}_1^K$ of K vectors $\mathbf{t}_k$ of lengths 1 through K where each $||\mathbf{t}_k|| > 0$. Note that elements of $\mathbb{T}^k$ can also be represented as rank k lower triangular matrices which is reflected below in the notation $\mathbf{T}$ used for elements of $\mathbb{T}^k$.

If using this algorithm to randomly generate matrices from $\mathbb{Q}$ or using it in a stochastic framework or gradient descent, then it needs to be imposed that for all $\mathbf{t}$ its magnitude $||\mathbf{t}_k|| \sim \mathcal{D}$, where $\mathcal{D}$ is a univariate continuous distribution with support $\mathbb{R}^+$, to ensure that all $\mathbf{t}_k$ are sampled uniformly in the radial domain. The author suggests to use the Gamma

distribution with shape parameter $\phi + 1$ and scale parameter $\frac{1}{\phi}$ where ϕ is the golden ratio, i.e. mean and standard deviation of one. A mean of one is desirable as when transforming from $\mathbb{Q}$ into $\mathbb{T}$ all $||\mathbf{t}||$ are of length one. Furthermore, this distribution has a density of zero for zero vectors which are invalid $\mathbf{t}$ values.

As the re-parametrisation is easier to understand from $\mathbb{T}^K$ to $\mathbb{Q}^K$, this direction is presented first. For each vector $\mathbf{t} \in \mathbf{T} \in \mathbb{T}^K$ calculate a unit vector $\mathbf{w}$ through division by its magnitude

$$\mathbf{w} = \frac{\mathbf{t}}{||\mathbf{t}||} \tag{3.21}$$

Then map $\mathbf{w}$ smoothly to a $\mathbf{v}$ which is guaranteed to be on one specific half of the unit hypersphere

$$\mathbf{v} = \frac{\mathbf{w} - \mathbf{e}_1}{||\mathbf{w} - \mathbf{e}_1||} \tag{3.22}$$

where $\mathbf{e}_1$ is a constant unit vector with 1 in the first dimension. This is equivalent to letting $\mathbf{v}$ be the reflection plane of $\mathbf{w}$ which reflects $\mathbf{w}$ through the origin. Note that there is one singularity here which is removable in the two dimensional case but becomes non-removable in higher dimensions. This singularity occurs when $\mathbf{w}$ is $\mathbf{e}_1$. This singularity occupies a part of the domain of permutation matrices. As the singularity ought to not cause issues when traversing it during gradient updates due to its limits symmetries and occupies only one "line" independent of the dimensionality of $\mathbf{w}$, thus zero density, in the parameter space it should be safe to ignore it. In the context of a computer algorithm, to avoid exceptions being thrown, treating $\mathbf{v}$ as $\mathbf{e}_1$ when $\mathbf{w}$ is $\mathbf{e}_1$ will still give a consistent and unique result in the next step. Alternatively treating $\mathbf{v}$ as a zero vector is simpler to implement, but the determinant and, thus, type of the resulting $\mathbf{Q}$ is negated.

From each $\mathbf{v}$ construct a householder matrix $\mathbf{P}$

$$\mathbf{P}_k = \mathbf{I} - 2 \cdot \mathbf{v}_k \mathbf{v}_k' \tag{3.23}$$

Then extend each such $\mathbf{P}$ to be of rank K by pre-pending 1 on the diagonal and 0 elsewhere then construct a rank K $\mathbf{Q}$ by taking the product over all $\mathbf{P}$

$$\mathbf{Q}_K = \prod_{k=K}^{1} \mathbf{I}_K + \mathbf{I}_K[:, k{:}K] \cdot (\mathbf{P}_k - \mathbf{I}_k) \cdot \mathbf{I}_K[k{:}K, :] \tag{3.24}$$

This $\mathbf{Q} \in \mathbb{Q}^K$ is uniquely defined for each $\mathbf{T} \in \mathbb{T}^K$.

In the inverse direction, it is possible to recover up to the $\mathbf{v}$s found in equation 3.21 from which a valid $\mathbf{T}$ can be formed

$$\mathbf{w}_k = \mathbf{Q}_k[:, 1] \tag{3.25}$$

Find the Householder matrix $\mathbf{P}_k$ as in equations 3.22 through 3.23 and then calculate

$$\mathbf{Q}_{k-1} = (\mathbf{P}_k \mathbf{Q}_k)[2{:}, 2{:}] \tag{3.26}$$

and repeat this steps until k becomes 1.

Note that each such calculated $\mathbf{w}_k$ is a unit vector.

Notes

The resulting $\mathbf{Q}$ is a rotation matrix, whose determinant is always $+1$, if its rank K is odd and $\mathbf{t}_1$ is positive or if K is even and $\mathbf{t}_1$ is negative. Otherwise, $\mathbf{Q}$ is a rotoreflection matrix, whose determinant is always -1. This follows from the algorithm as $\frac{\mathbf{t}_1}{||\mathbf{t}_1||} = \mathbf{Q}_1$ and each subsequent increase of rank is done through a householder reflection. The initial 1 by 1 matrix -1 is a 1D reflection and 1 a rotation of 0 deg and two reflections negate each other. Thus if it is desirable to restrict a $\mathbf{Q}$ to either only rotation or only rotoreflections, one can do so by fixing $\mathbf{t}_1$ to 1 or -1 respectively depending on the rank of $\mathbf{Q}$.

Implementation

There are a number of optimisations which can be performed while implementing this algorithm. Firstly $\mathbf{Q}$ can be initialised as an identity matrix of rank K and updated in place. Secondly transforming $\mathbf{t}_k$ to $\mathbf{P}_k$ can be folded into updating $\mathbf{Q}$ in a way which omits multiple of the intermediary steps.

$$\mathbf{Q} = \mathbf{I}_K \tag{3.27}$$

For each $\mathbf{t} \in \mathbf{T} \in \mathbb{T}^K$

$$n_t = ||\mathbf{t}|| \tag{3.28}$$

if $\mathbf{t}[1] \neq n_t$

$$\mathbf{t}[1] \mathrel{-}= n_t \tag{3.29}$$

$$\mathbf{Q}[k{:}K, k{:}K] \mathrel{+}= \frac{\mathbf{t}\mathbf{t}'\mathbf{Q}[k{:}K, k{:}K]}{n_t \mathbf{t}[1]} \tag{3.30}$$

else, premultiply $\mathbf{Q}$ by a diagonal matrix of all 1 except the bottom most which is -1.

After the loop $\mathbf{Q}$'s value is the desired output of the algorithm. If the algorithm is to be used in a stochastic framework or for gradient descent, then a penalty p can be accumulated easily within the loop by splicing the following operation into each iteration

$$p \mathrel{+}= \text{logpdf}_{\mathcal{D}}(n_t) \tag{3.31}$$

with p initialised to 0 at the beginning. p can then at the end be combined with the loss or log posterior as appropriate for the algorithm used.

Note that this implementation will threat any $\mathbf{t}$ of length zero, which are not valid $\mathbf{t}$s and have a density of zero in the proposed penalty $\mathcal{D}$ gamma distribution, the same as $\mathbf{e}_1$.

Parts of the implementation of the inverse direction can be similarly simplified.

For $k \in \{K, ..., 1\}$ if $k \neq 1$ do

$$\mathbf{t}_k = \mathbf{Q}_k[:, 1] \tag{3.32}$$

if $\mathbf{t}_k[1] \neq 1$ do the following steps

$$\mathbf{t} = \mathbf{t}_k \tag{3.33}$$

$$\mathbf{t}[1] \mathrel{-}= 1 \tag{3.34}$$

$$\mathbf{Q}_k \mathrel{-}= \frac{\mathbf{t}\mathbf{t}'\mathbf{Q}_k}{1 - \mathbf{t}_k[1]} \tag{3.35}$$

$$\mathbf{Q}_{k-1} = \mathbf{Q}_k[2:, 2:] \tag{3.36}$$

$\mathbb{Q}$ to $\mathbb{T}$ to $\mathbb{Q}$ always round-trips up to floating point error. $\mathbb{T}$ to $\mathbb{Q}$ to $\mathbb{T}$ only does so up to the scale of each $\mathbf{t}$ and all zero vector $\mathbf{t}$ are replaced with a $\mathbf{e}_1$ vector.

3.6 Maximum A Posteriori ICA

To inject prior information into ICA, using a Bayesian framework, a likelihood function $\mathcal{L}$ as described in section 3.3 for the ICA part and a prior π need to be defined as per the common Bayesian equation:

$$\Pi(\mathbf{W}|\mathbf{X}) \propto \mathcal{L}(\mathbf{W}; \mathbf{X})\pi(\mathbf{W}) \tag{3.37}$$

and in the computationally efficient log domain:

$$\log \Pi(\mathbf{W}|\mathbf{X}) = \log \mathcal{L}(\mathbf{W}; \mathbf{X}) + \log \pi(\mathbf{W}) + c \tag{3.38}$$

where c is the log of the proportionality constant before the transform.

It is however very difficult to determine a meaningful prior distribution of $\mathbf{W}$ thus a prior on each column of its inverse $\mathbf{M}$ is used instead. The procedure to derive said prior distribution $\mathcal{N}_\mathcal{M}$ on each column of $\mathbf{M}$ is described earlier in section 3.4. Given that distribution, the log-prior of $\mathbf{M}$ is

$$\log \pi(\mathbf{M}) = 2||\mathbf{M}|| + \sum_{s}^{S} \log p_{\mathcal{N}_\mathcal{M}}(\mathbf{M}[:, s]) \tag{3.39}$$

the factor before the sum accounts for the prior being on $\mathbf{M}$ and not on $\mathbf{W}$. Without enforcing $\mathbf{W}_1$ to be orthogonal the posterior would then be

$$\log \Pi(\mathbf{W}|\mathbf{X}) = \log \mathcal{L}(\mathbf{W}_1; \mathbf{X}_0) + \log \pi(\mathbf{M}_0\mathbf{W}_1^{-1}) + c \tag{3.40}$$

which contains an inverse prone to cause numerical instabilities during gradient descent.

Provided the orthogonality constraint on $\mathbf{W}_1$ is maintained the posterior becomes

$$\log \Pi(\mathbf{W}|\mathbf{X}) = \log \mathcal{L}(\mathbf{W}_1; \mathbf{X}_0) + \log \pi(\mathbf{M}_0\mathbf{W}_1') + c \tag{3.41}$$

which is much less prone to numerical issues. Using the re-parametrisation defined in section 3.5 on $\mathbf{W}_1$ gives this loss function:

$$\text{loss}(\mathbf{T}|\mathbf{X}_0) = -\left(\log \Pi(\text{TtoQ}_{\mathbf{W}_1}(\mathbf{T})|\mathbf{X}_0) + \text{TtoQ}_{\text{penalty}}(\mathbf{T})\right) \tag{3.42}$$

which can be optimised by any gradient descent algorithm although those that are able to adjust learning rates for each value in $\mathbf{T}$ independently are likely to give better results as the importance of elements in $\mathbf{T}$ is unlikely to be evenly distributed.

3.7 Full Posterior

If the penalty term described in section 3.5 is added, then the above log-posterior function can be used in a stochastic framework, such as Stan (see appendix B), to acquire samples from the posterior distribution from which an approximation of the full posterior can be derived. Both Markov Chain Monte Carlo and Variational Bayes methods are feasible on modern hardware if only a few experiments are planned, but the time requirements can get onerous if many separate experiments are executed.

Instead, the Laplace approximation [125] can be used if the trade-off between accuracy and time is considered acceptable. To calculate the Laplace approximation $\tilde{\mathbf{W}}$ of the full-posterior, first the maximum a posteriori $\hat{\mathbf{W}}$ must be estimated using equation 3.42, then, after taking the Hessians $\hat{\mathbf{H}}_s$ of the loss function with regards to each row $\hat{\mathbf{w}}_s$ of $\hat{\mathbf{W}}$, where s is the row index, the approximation is:

$$\begin{aligned} \hat{\Sigma}_s &= (-\hat{\mathbf{H}}_s)^{-1} \\ \hat{\mu}_s &= \hat{\mathbf{w}}_s \end{aligned} \tag{3.43}$$

The sought $\tilde{\mathbf{W}}$ is the combination of all such estimated $\hat{\Sigma}_s$ and $\hat{\mu}_s$ for all s, that is for each sample t in each row in $\mathbf{W}$:

$$\mathbf{w}_s \,|\, \mathbf{X} \sim \mathcal{N}(\hat{\mathbf{W}}[t,:], \hat{\mathbf{\Sigma}}_s) \tag{3.44}$$

It is worth of note that the estimated $\hat{\Sigma}_s$ may be underestimated due to effects of the autocorrelation in $\mathbf{Y}$ and thus $\mathbf{X}$ as common in time-series data. This can be alleviated by downsampling $\mathbf{X}$ in the time dimension; however, care needs to be taken as ICA becomes unreliable if the frequency content of the sought signals approaches too close, to the Nyquist frequency. Please see section 4.2 for the downsampling parameters used in the experiments run as part of this thesis.

3.8 Predictive Distributions

Given EMG observations $\check{\mathbf{X}}$ assumed to be generated by some unknown sources $\mathbf{Y}$ through a mixing process $\hat{\mathbf{W}}$ as estimated from $\mathbf{X}$ which may or may not also be $\check{\mathbf{X}}$, the posterior predictive distribution of $\tilde{\mathbf{Y}}$ can be calculated as follows:

$$\tilde{\mathbf{Y}}[s,t] \,|\, \mathbf{X}, \check{\mathbf{X}} \sim \mathcal{N}(\hat{\mathbf{W}}[s,:]\check{\mathbf{X}}[:,t], \check{\mathbf{X}}[:,t]'\hat{\mathbf{\Sigma}}_t\check{\mathbf{X}}[:,t]) \tag{3.45}$$

In particular, the mean and standard deviation of the sth source at the tth sample of the signal distribution $\tilde{\mathbf{Y}}$ can be found as:

$$\begin{aligned} \mu_{st} &= \hat{\mathbf{W}}[s,:]\check{\mathbf{X}}[:,t] \\ \sigma_{st} &= \sqrt{\check{\mathbf{X}}[:,t]'\hat{\mathbf{\Sigma}}_f\check{\mathbf{X}}[:,t]}. \end{aligned} \tag{3.46}$$

A visual representation of such a $\tilde{\mathbf{Y}}$ can be found in figure 4.3.

The maximum likelihood predictive distribution can be analogously calculated through a "posterior" of $\mathbf{W}$ with a flat prior respectively instead of $\tilde{\mathbf{W}}$.

As the standard deviation is sensitive to the ambiguous scaling produced by ICA, a normalised score which can be understood as a signal to noise ratio of the posterior predictive distribution is calculated for each source s in $\tilde{\mathbf{Y}}$:

$$\mathrm{SNR}[s] = \frac{\sum_t \boldsymbol{\mu}^2[s,t]}{\sum_t \boldsymbol{\sigma}^2[s,t]} \tag{3.47}$$

SNR here is calculated as the ratio of the total power of $\boldsymbol{\mu}_s$, which stands in as an approximation of the true signal, and the total power of the deviations from that signal.

Note that this signal to noise ratio is an abstract proxy of the certainty of the sources, its value is not directly related to the ratio of "true signal" and "noise" of the unmixed sources. At the limit, this SNR becomes ∞ in the case of absolute certainty with $\sigma = 0$, and 0 in the case of absolute uncertainty with $\sigma = \infty$.

3.9 Summary

This chapter presented a framework to calculate the predictive posterior distribution of signals estimated using independent component analysis as well as an approach to analyse the latter to assign a quantitative descriptor of the certainty one can have in these results. Also described was an approach to estimate a prior probability function for mixture matrices based on a forward model. As the variable in ICA is the un-mixing matrix rather than the mixing matrix, this chapter also demonstrated how to handle the inversion during the calculation without suffering from catastrophic floating-point errors by introducing a reparametrisation of orthogonal matrices which allowed that inverse to become a simple transpose while calculating the prior probability.

4 Experimental Results

This chapter will briefly present the prior used, some real and synthetic EMG datasets. Then it will continue to describe and discuss the results of applying the algorithms defined in the previous chapter to them.

The results from the synthetic data are presented first. They are used to establish a baseline as to what to expect given certain qualities in a given dataset. Then, the results of the real EMG datasets are used to demonstrate that the expectations created by the synthetic data hold in the real world.

4.1 Generated Prior

The recording electrode grid is assumed to be a 9 by 9 square with the distance between the centre of each electrode being equal to 5mm. This is based on the actual electrodes used to gather the data.

MUAPs are assumed to be uniformly distributed parallel to the skin in an area defined by the outermost electrodes. The depth of them is assumed to be normally distributed with a mean of 3.68 mm below the fat layer and a standard deviation of 1.56 mm. A graphical presentation of this setup is found in figure 4.1.

To prevent MUAP locations within the fat layer, the depth distribution is approximated by sampling $z_{\mathcal{N}} \sim \mathcal{N}$ and then applying the following transformation: $\sqrt{|20 \cdot z_{\mathcal{N}}|}$, which returns a depth value in millimetre with approximately the above described distribution. This approximation is arbitrary, but is not expected to affect the results noticeably, particularly since the choice of the underlying MUAP distribution is already based on an educated guess rather than the result of hard evidence.

Given the model as described in section 2.2 and the above MUAP location distributions, any number of sample rows of the mixing matrix can be generated. For the synthetic datasets described below, one mixing matrix was generated by stacking 81, as there are also 81 electrodes, randomly generated rows. For the prior to be used to find the respective posterior distributions, 10000000 column vectors were randomly generated and then a multivariate normal was fitted over these column vector samples. This amount of samples was imposed by the limitations of the hardware used to perform the calculations. Less samples would also have been most likely sufficient.

Model Comparison

Figure 4.2 shows a comparison to the result from the cylindrical finite element model in [79]. Particularly, this figure is setup to match figure 11 there, which shows the normalised signal

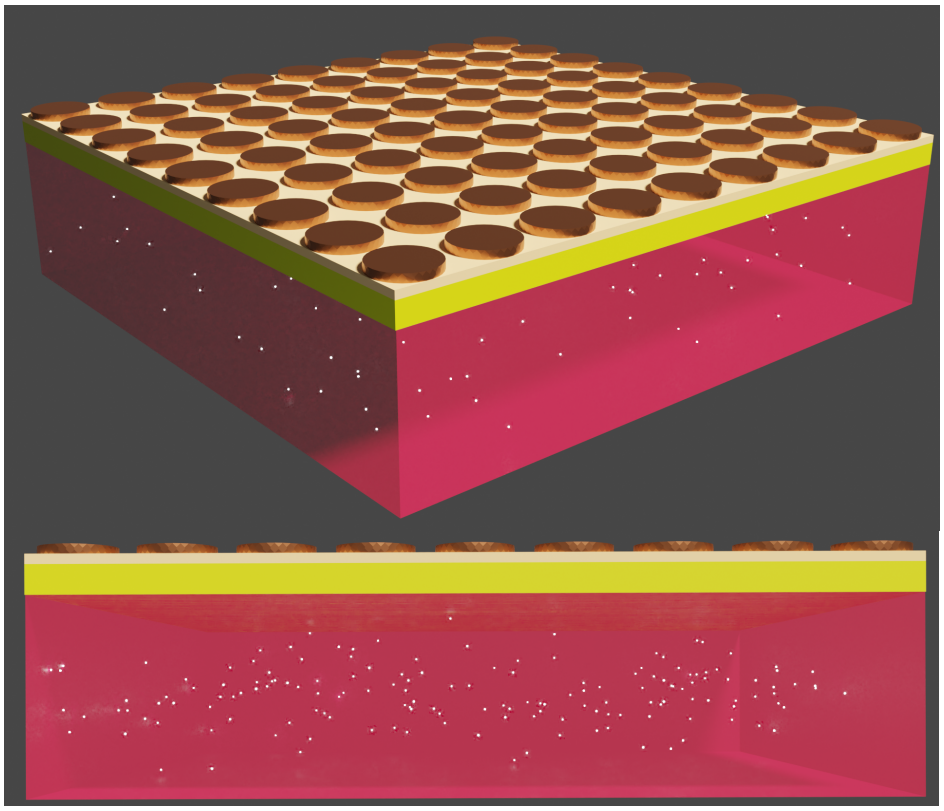

Figure 4.1: This figure shows a to-scale model of the layers and electrode configuration as assumed when calculating a prior on $\mathbf{M}$ in section 4.1. The white dots represent MUs but their number was adjusted to show more clearly their distribution inside the muscle.

value at an electrode given a specific angular displacement from the emitting source. As expected, there is some difference, both because the large difference in model complexity and because of the transform needed to compare a planar and a cylindrical model. The author considers this agreement to be sufficient given those circumstances to deem the implemented model valid.

4.2 Experimental Setup

Synthetic Datasets

As an approximation, the shape of real MUAPs can be considered similar to that of a Mexican hat wavelet which itself can be approximated as the difference of two Gaussians. From such spike templates synthetic MUAP trains can be generated given a spike interval. These intervals between the spikes were chosen to be lognormally distributed. The simulated sampling frequency was 2048 Hz with a length of 10240 samples, i.e 5 seconds, to match the real datasets.

As it is expected that in the estimated $\mathbf{Y}$ of a real dataset only some "sources" represent MUAP traces with the rest being background noise from MUs located too deep to recognise,

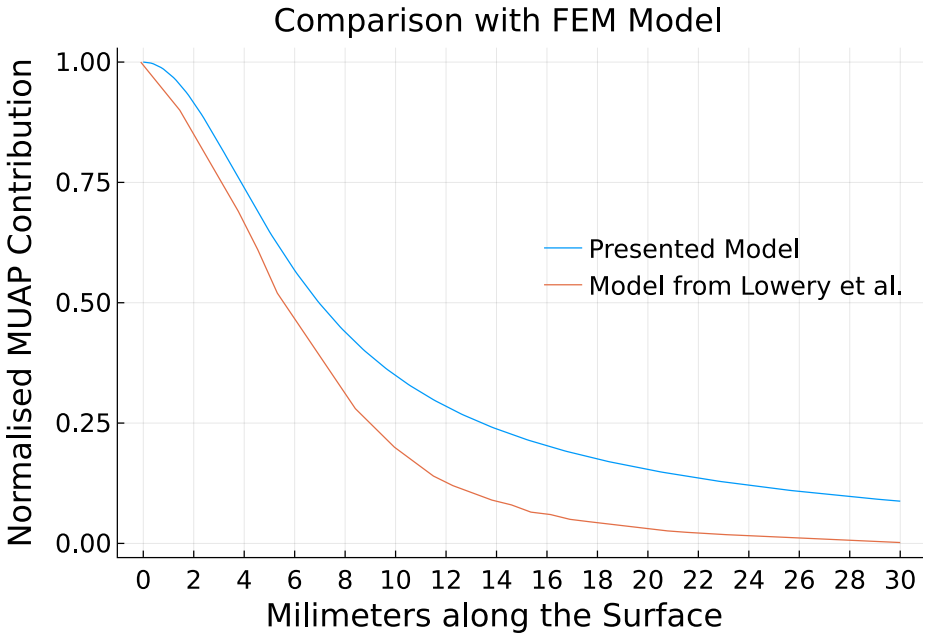

Figure 4.2: This graph is setup to allow comparison with figure 11 in [79]. To this end, the values from that graphs x axis have been transformed such as they would have been approximately in a planar geometry.

multiple datasets were generated with a variable number of generated MUAP traces. Specifically, each generated synthetic $\mathbf{Y}$ contains 81 "sources" of which 9, 18, 27, 36, 45, 54, 63, 72 or 81 are generated MUAP traces with the remaining "sources" being filled with Gaussian noise to model the background signals of MUs located too deep to result in recognisable MUAP traces.

The synthetic observation $\mathbf{X}$ is calculated for each synthetic $\mathbf{Y}$ by multiplying that $\mathbf{Y}$ with a mixing matrix $\mathbf{M}$ sampled from a physiological model as described above.

The observation noise found in real datasets was then simulated through an even mixture of Gaussian and Laplacian noise which was added to $\mathbf{X}$ after mixing. Multiple noise magnitudes were added to observe the effects of different signal to noise ratios of the observations. In particular, the following magnitudes of noise relative to the magnitude of the data were added: 0.5, 0.75, 1, 1.25, 1.5, 2, 3, 4, 5, 6, 8.

Both the ranges for the amounts of real sources and amounts of noise where determined by preliminary experiments. For other datasets, the interesting areas of the parameter space may lie elsewhere.

DeTOP Datasets

As real data, this paper uses the datasets from [141], where the electrodes are affixed to the forearm while the subjects performed various hand movements as described in [142], which is an extensive collection of EMG recordings. They are at a 2048 Hz sampling frequency.

The datasets contain 9-times-14 observation channels but only the top 9-times-9 square of these are used here.

There are generally periods of inactivity at the beginning and end of each of the DeTOP datasets. Thus, from each dataset, the continuous block of 10240 samples (that is 5 seconds), with the highest mean absolute magnitude, was assumed to be the block of most interest. The validity of that selection, i.e. that the selection is not caused by artefacts, was inspected visually for each dataset used.

Pre-processing

A number of preparatory steps are common to both the synthetic and real data.

EMG data is highly correlated in the time dimension, particularly at high sampling rates, i.e the 2048 Hz the used date is in. To alleviate this concern, each dataset was downsampled to 512 Hz using a digital filter, that is to 2560 samples each dataset.

Each of these smaller datasets was then further split into a training and a test set. That is, only one part, the training part, is used to derive a posterior distribution, the remaining testing part is only used during analysis at the end. This allows to ward against false conclusions due to potential over-fitting. As the effect of different training set sizes is of interest, the experiment was rerun with a number of different training set sizes. The different sizes for a training set where: 512, 770, 1024, 1280, 1536, 1792, 2048. The training data is always the centremost while the test set is composed on the data remaining before and after the training data.

There may be a small amount of information leakage between the test and training parts of each dataset at the borders between them, however, this is deemed negligible as there are only a few affected samples which may be correlated, particularly as the downsampling already intends to minimise any correlation between adjacent samples. In addition to this, it should be kept in mind that the noise is most likely of identical nature in both the training and test parts of a given dataset.

4.3 Examination of Results From Synthetic Datasets

As noted before, the experiment has been run over all the DeTOP datasets. The results from each run have been found on visual inspection to be very similar, across all of them, to the subset shown in this section.

Posterior Predictive Distribution

Each source in $\mathbf{Y}$ has its own posterior predictive distribution, but due to ICA scaling ambiguity, they are not directly comparable. To compare them, one needs to calculate a measure which is not sensitive to the scaling such as signal to noise ratio discussed in previous

chapters or to rescale them based on some heuristic. The latter can be useful to allow manual qualitative comparison of the same source retrieved from differently parametrised executions of ICA.

As discussed in section 4.2, an experiment was run using synthetic data using different levels of noise of the observations, different amounts of samples and either a informative or non-informative prior.

To match the estimated sources to the ground truth in figures 4.3, 4.5 and 4.6 the Gale–Shapley algorithm was used with absolute correlation as heuristic described in section 2.3, then, for figure 4.3, the chosen sources had their magnitude normalised to the magnitude of the spike of interest in the ground truth.

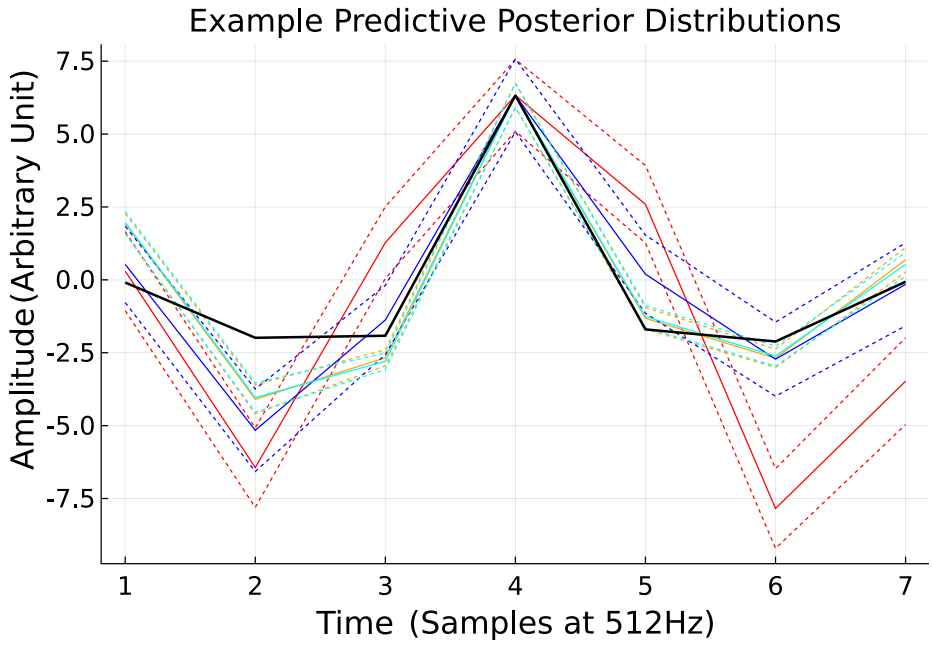

Figure 4.3: The bold black line is the truth as determined by correlation matching, the solid lines are the estimated signal means and the dashed lines represent each time one standard deviation above and below the mean. The red source has been recovered from an observation with low SNR without using prior information. The blue source has been recovered from an observation with low SNR while using prior information. The orange source has been recovered from an observation with high SNR without using prior information. The cyan source has been recovered from an observation with high SNR while using prior information. All sources have been rescaled to match the magnitude of the ground truth as ICA does not recover scale information.

Figure 4.3 shows an illustrative example of the differences between the posterior predictive distributions of one source as resulting from different experiment parameters. As can be seen in the figure, the level of noise of the input observations has a large effect independent of the informativeness of the prior, while an informative prior only visibly improves the result in the case where the observation is noisy. This pattern of the quality of the result and of the

effect of the prior given the observations signal to noise level is consistent across datasets and measurements.

Effect of the Prior

Taking the difference between the SNR, averaged over all features, between runs with an informative prior and runs without gives further objective evidence to that subjective assessment. As visualised in figure 4.4, The average SNR is always positively affected by an

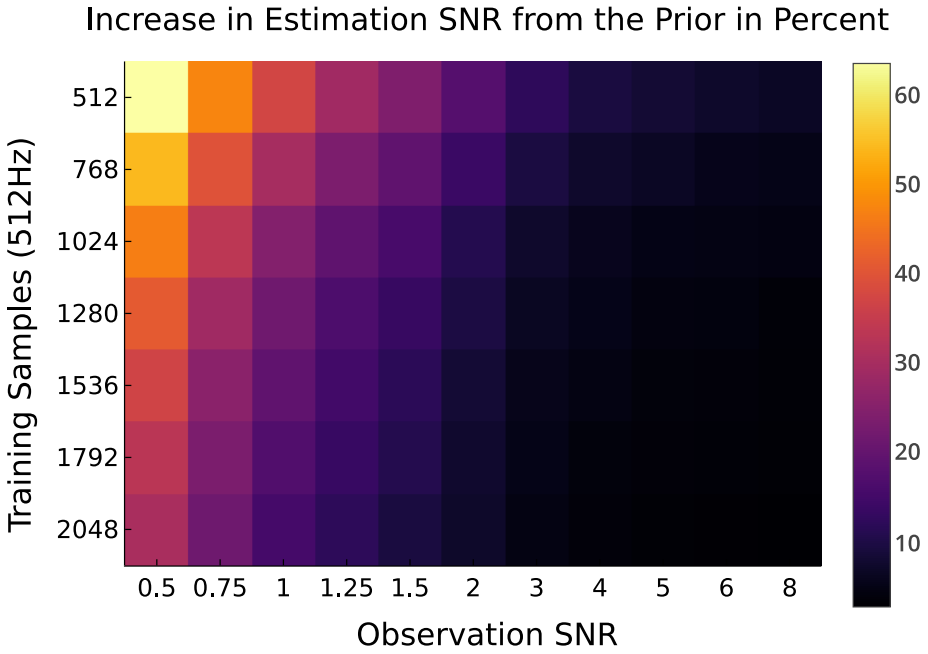

Figure 4.4: This heat map shows the increase of SNR of the recovered signal given observation SNR and training data samples averaged over all sources and time points.

informative prior but much more so when there are few input samples or high observation noise or both. It is to note that a doubling of the observation noise has a much more significant effect on the usefulness of an informative prior than a halving of the samples in the input observations.

The effect of an informative prior does not seem always homogeneous over all features. If looking at the correlation between the estimated features and the ground truth, there often isn't a visible difference when using an informative prior and when not which does not lie well within the general variance observed within these results. Where it differs, however, it principally does as seen in figure 4.5 where, for the sources with the highest correlation, the informative prior is deleterious, but beneficial for the less well recovered sources and no difference exists for white noise pseudo-sources.

Figure 4.6 shows the full correlation matrices. The seemingly high correlation with the pure noise sources 28 and later is the result of correlation with noise in-between the MUAP

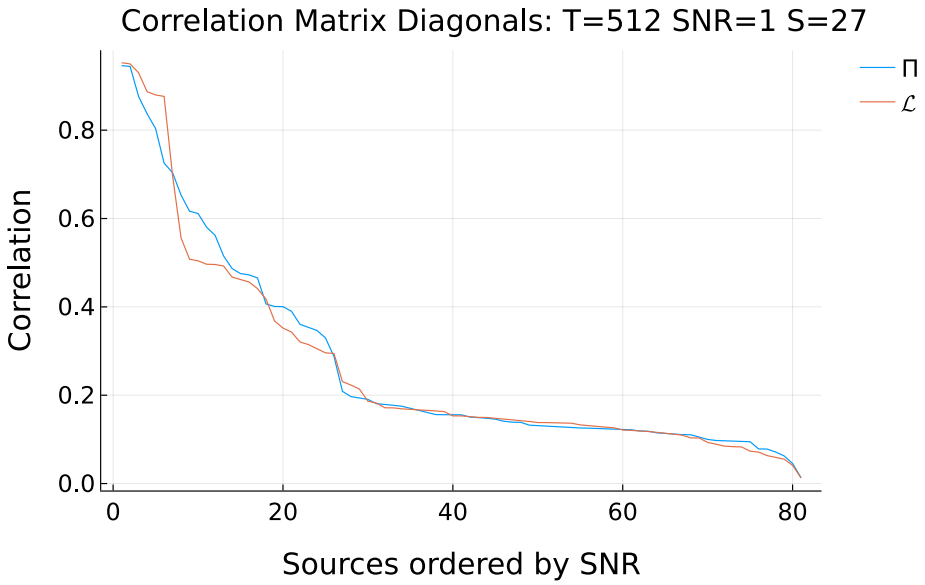

Figure 4.5: The diagonals of the correlation matrices between the ground truth and the estimates with and without an informative prior. There are 27 actual sources in the ground truth, the rest is white noise simulating the background noise. Both experiments were run using 512 training samples per source and noise of equal magnitude to the observation was added to the observation after mixing.

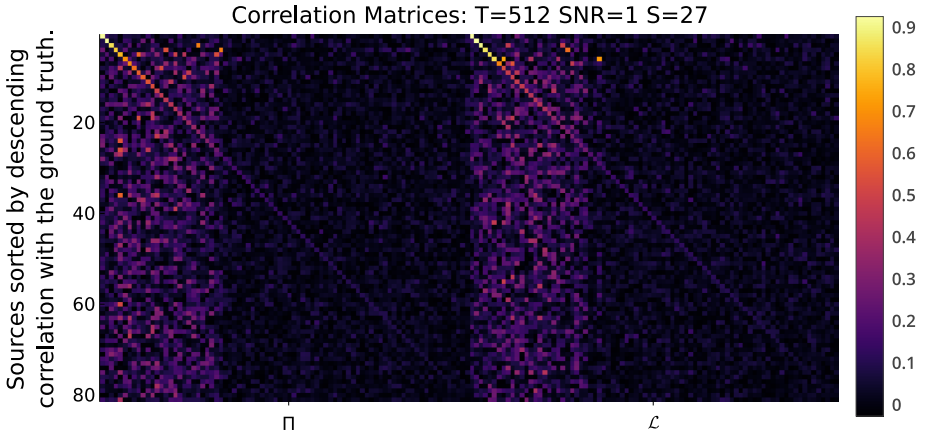

Figure 4.6: The correlation matrices between the ground truth and the estimates with and without an informative prior. There are 27 actual sources in the ground truth, the rest is white noise simulating the background noise. Both experiments were run using 512 training samples per source and noise of equal magnitude to the observation was added to the observation after mixing.

events, when only taking the correlation for the parts where there are MUAP events, then the contrast between high and low correlation in the matrices is much higher.

While it is hard to draw objective conclusions from figure 4.6, it can be qualitatively observed that adding an informative prior decreases SNR at the sources with the most SNR but increases SNR for all other non-noise sources. This would be in line with the expectation

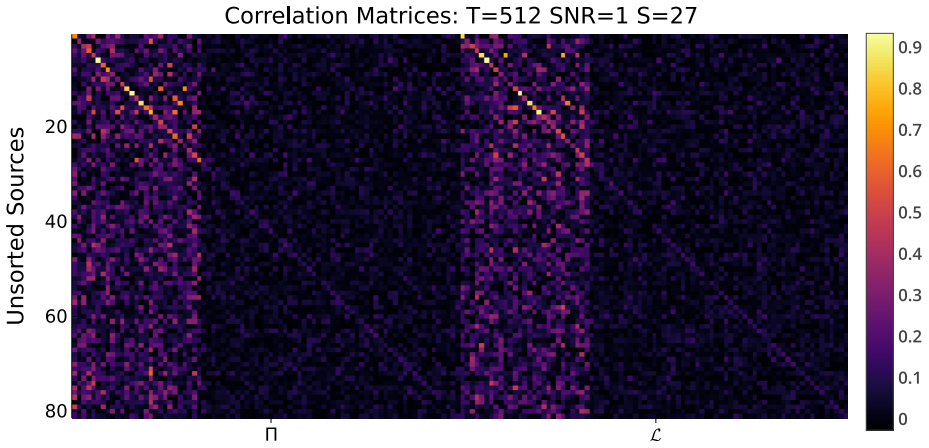

Figure 4.7: The same data as in figure 4.6 but before sorting the sources.

one would have if it is assumed that adding a prior reduces variance at the cost of adding bias which is the usual effect of adding a prior.

For comparison, figure 4.7 shows the same data before the sources were matched. As it is expected, they not ordered in any consistent form and of little use in this way.

Signal to Noise Ratios

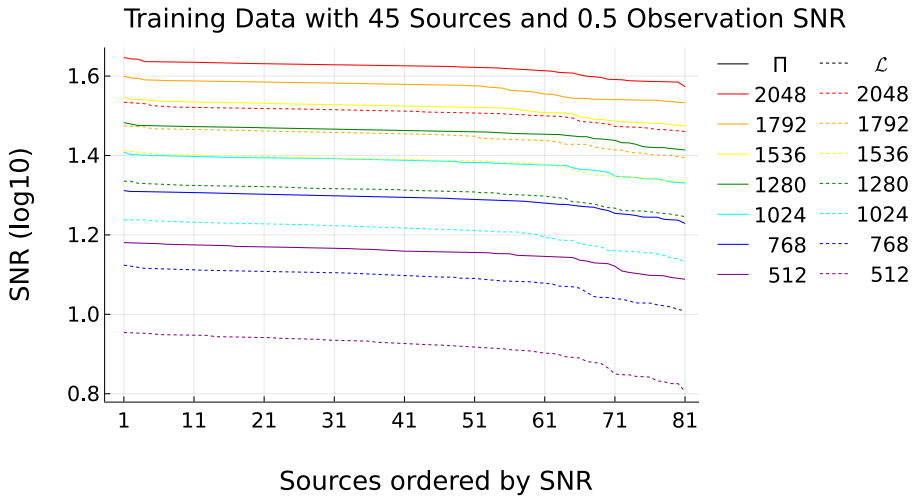

Figure 4.8: Signal to noise ratios for all estimated sources ordered by SNR estimated for various sample sizes from a very noisy synthetic observation dataset. The solid lines were calculated with a prior, the dashed ones without. The calculation was made using the same data as was used to estimate $\mathbf{W}$.

Figures 4.8 to 4.11 show the found signal to noise ratios by source for a range of training sample sizes both with and without informative prior. Figures 4.10 and 4.11 are the SNR calculated on the testing part of the dataset while figures 4.8 and 4.9 are based on the training

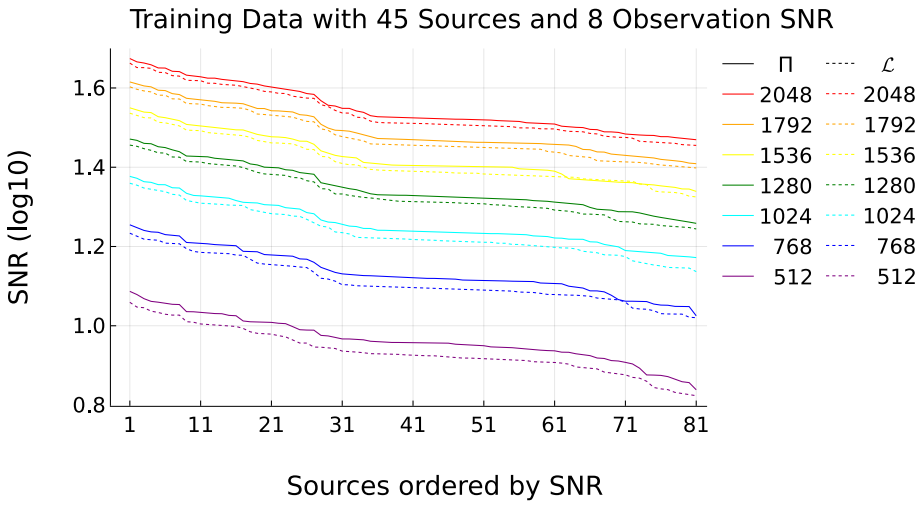

Figure 4.9: Signal to noise ratios for all estimated sources ordered by SNR estimated for various sample sizes from a synthetic observation dataset with only little added noise. The solid lines were calculated with a prior, the dashed ones without. The calculation was made using the same data as was used to estimate $\mathbf{W}$.

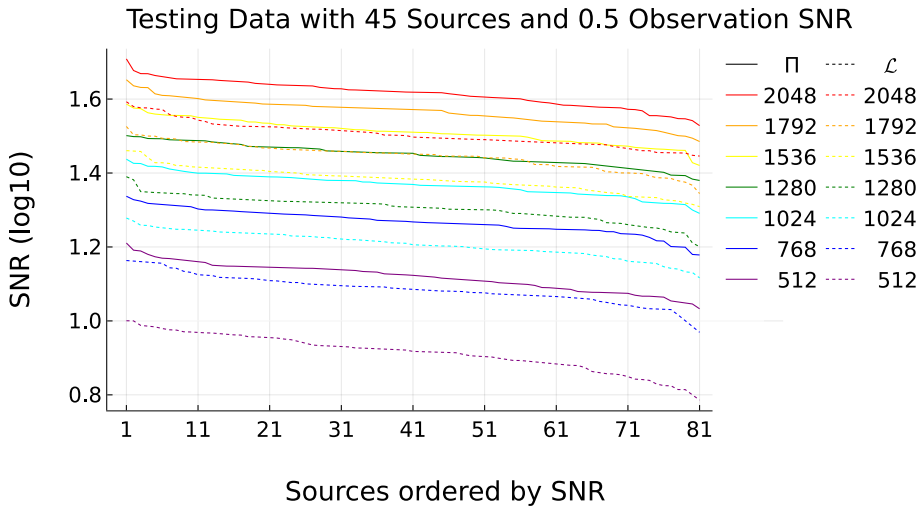

Figure 4.10: Signal to noise ratios for all estimated sources ordered by SNR estimated for various sample sizes from a very noisy synthetic observation dataset. The solid lines were calculated with a prior, the dashed ones without. The calculation was made using data separate from that used to estimate $\mathbf{W}$.

data itself. Similarly, figures 4.8 and 4.10 are calculated with noisy training observations while figures 4.9 and 4.11 are based on training data with very little noise.

As noted before, an informative prior has clearly a much more pronounced effect when there is a large amount of noise in the input data.

As ICA cannot differentiate actual sources from noise sources, it is not particularly surprising that the figures indicate that there is a higher SNR for noise sources when using

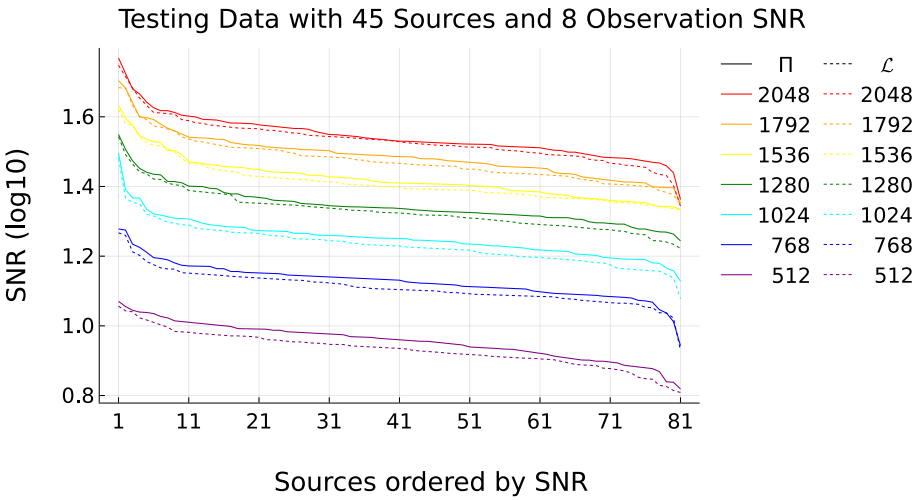

Figure 4.11: Signal to noise ratios for all estimated sources ordered by SNR estimated for various sample sizes from a synthetic observation dataset with only very little added noise. The solid lines were calculated with a prior, the dashed ones without. The calculation was made using data separate from that used to estimate $\mathbf{W}$.

a large training set than for actual sources when using few training samples. Thus, this SNR cannot be directly compared when the parameters of an experiment differ too much. However, it can be compared between source estimates of the same run as is clearly visible in figure 4.9. The sources before the sharp drop, slightly before source 30 for each configuration in that figure, have been visually confirmed to be the unmixed actual sources, although only partially in some cases, while the ones after the drop are mostly noise sources or very poorly unmixed sources depending on the other parameters. Similar transitions can be found in figure 4.8 and figure 4.10 around the third sources and around the roughly before the tenth sources in figure 4.11.

4.4 Examination of Results From Real Datasets

The results from the real datasets generally seem to support the deductions made from examining the results of the synthetic data experiment. For each of the below described results, the estimated source traces were visually inspected to confirm if their "unmixing" quality was as expected from their SNR.

Figures 4.12, 4.13, 4.14 and 4.15 represent good examples of datasets where only a few sources were unmixed well with a lot of sources which were essentially noise. The stark contrast between runs with many training samples and runs with few training samples is an artefact due to this dataset containing generally very sparse data and thus there genuinely were less sources to un-mix in the smaller training sets.

Figures 4.16, 4.17, Figures 4.18 and 4.19 are an example of cases where there are a few good estimate sources and many of mediocre quality. By mediocre quality is to be

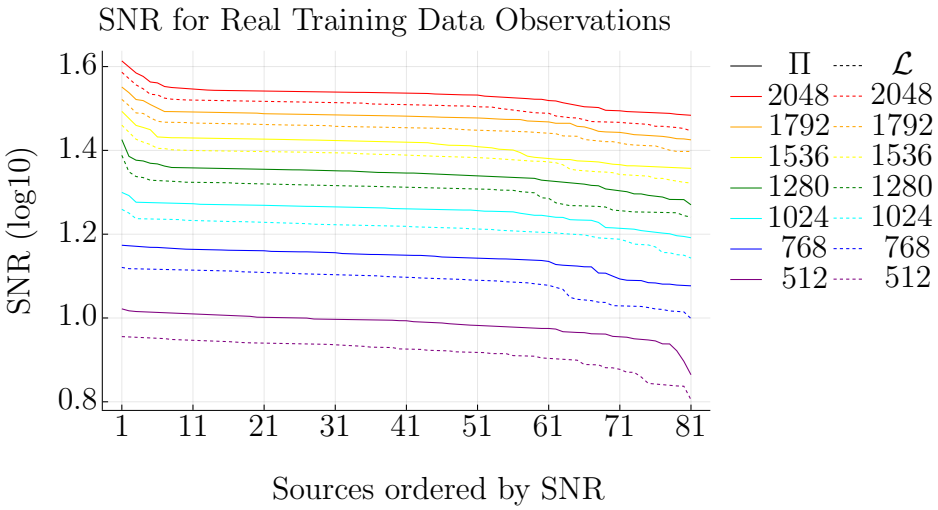

Figure 4.12: Signal to noise ratios for all estimated sources ordered by SNR estimated for various sample sizes from a observation dataset containing real EMG. The solid lines were calculated with a prior, the dashed ones without. The calculation was made using the same data that used to estimate $\mathbf{W}$.
The well unmixed sources correspond to the much stronger slope at the sources before source 10.

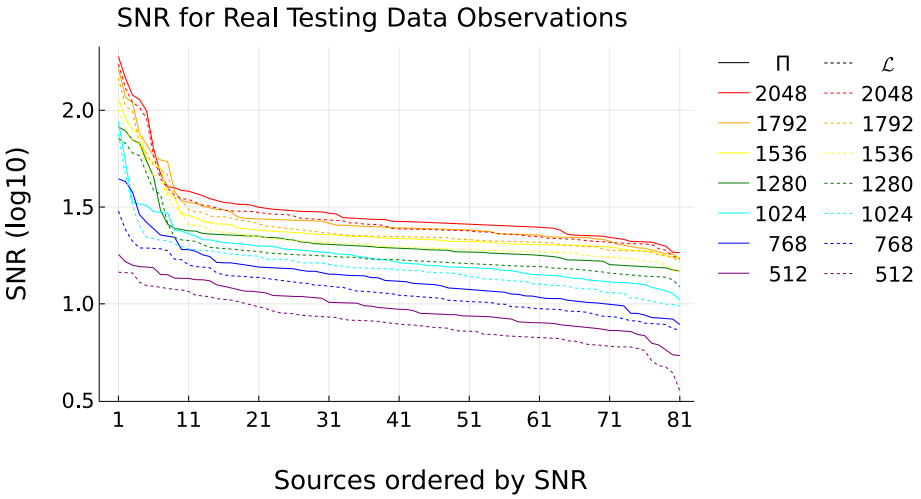

Figure 4.13: Signal to noise ratios for all estimated sources ordered by SNR estimated for various sample sizes from a observation dataset containing real EMG. The solid lines were calculated with a prior, the dashed ones without. The calculation was made using data separate from that used to estimate $\mathbf{W}$.
The well unmixed sources correspond to the much stronger slope at the sources before source 10.

understood that the spikes are recognisable if one is looking for them, but without knowing a priory the nature of the data this traces would be mistaken for noise. Also, particularly visible here is that the test data produces much more pronounced differences between real

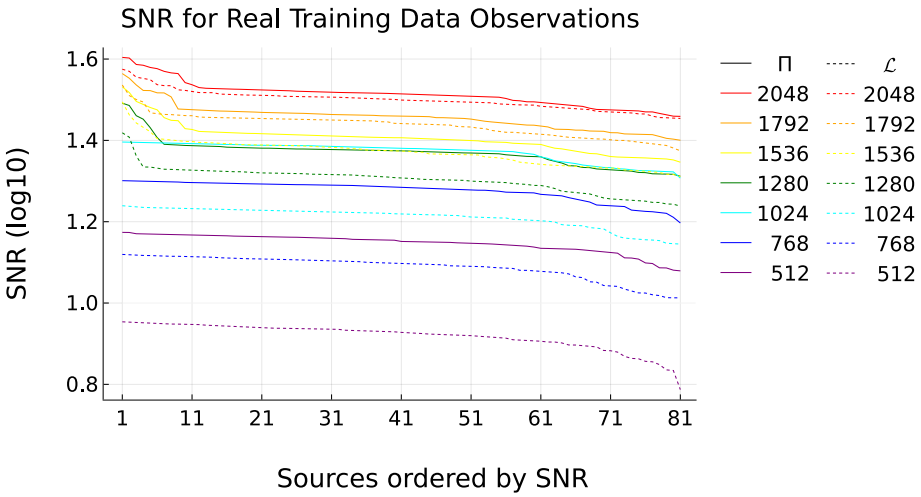

Figure 4.14: See figure 4.12 for a general description.
The well unmixed sources correspond to the much stronger slope at the sources before source 10 for runs with 1280 training samples or more.
The runs with less training samples only contain one or two visibly unmixed sources on the other hand.

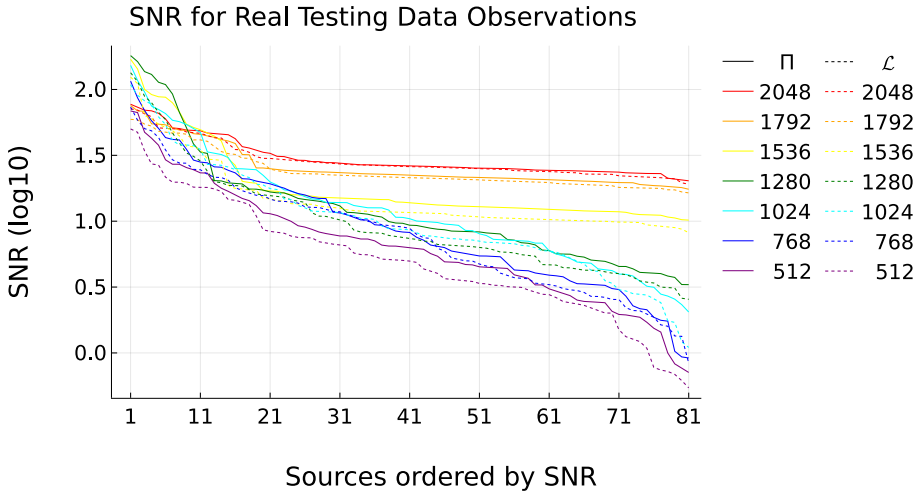

Figure 4.15: See figure 4.13 for a general description.
The same pattern as in figure 4.14 can be observed here, if in a more extreme fashion.

and noise sources than the training data.

Figures 4.20 and 4.21 serve as contrasting example to the above figures. None of the sources is unmixed well which results in no visible transition at the low numbered sources in the training data figure. The few mediocre sources are again more visible in the testing data figure then in the training data figure.

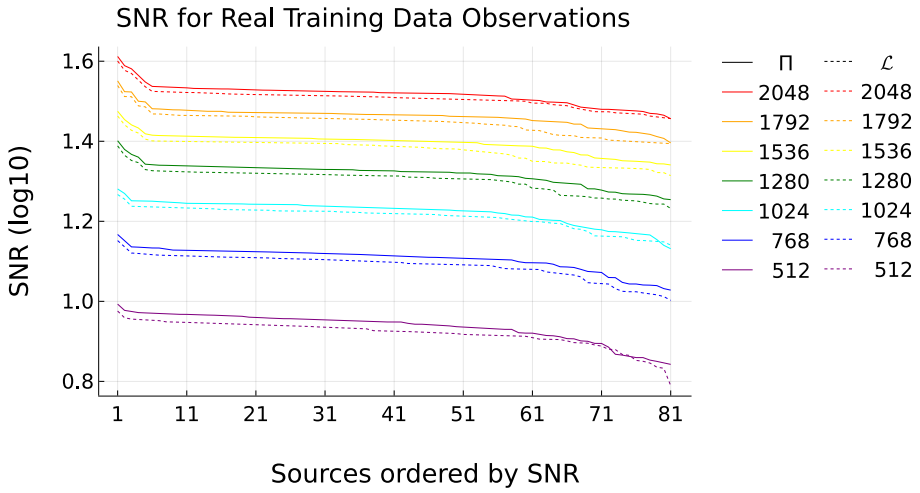

Figure 4.16: See figure 4.12 for a general description.
The slope at the very beginning corresponds to well unmixed sources. The plateau afterwards corresponds to 'mediocre' sources with steadily degrading un-mix quality. Where the slope accelerates downwards again, towards the end, the sources have become unrecognisable noise.

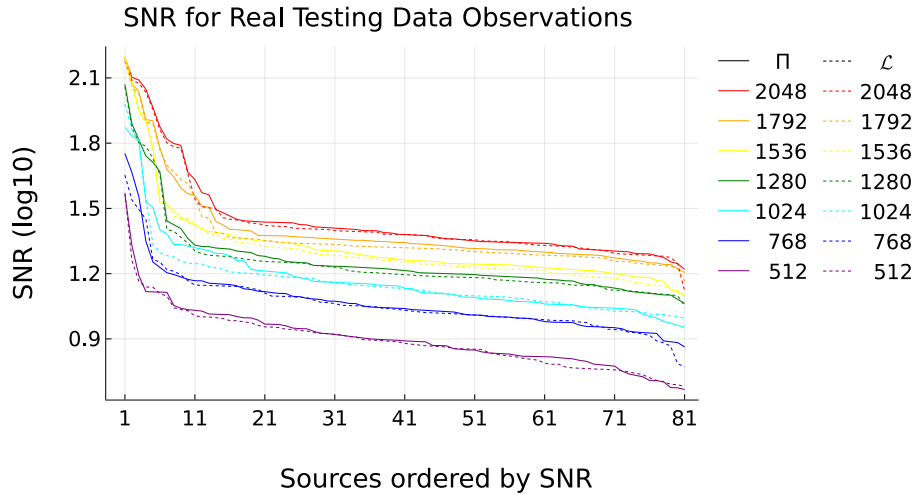

Figure 4.17: See figure 4.13 for a general description.
Same pattern as in figure 4.16 but the beginning is more pronounced.

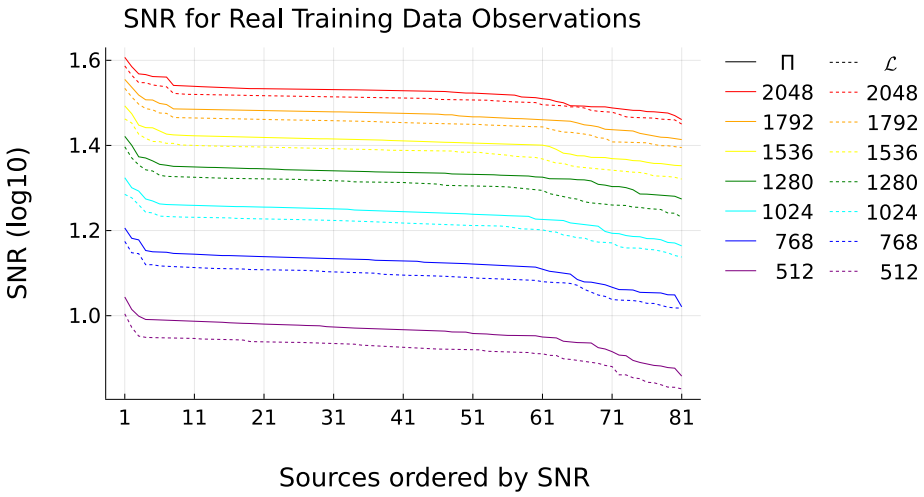

Figure 4.18: See figure 4.12 for a general description.
Same pattern as in figure 4.16.

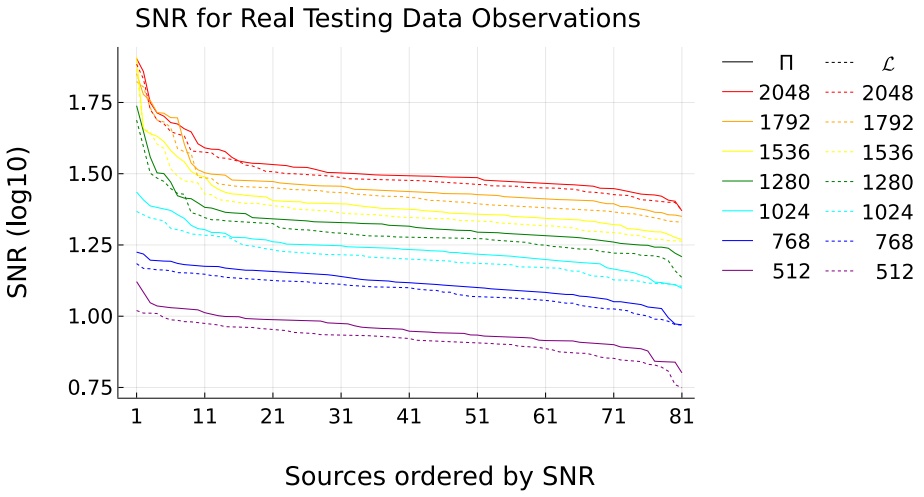

Figure 4.19: See figure 4.13 for a general description.
Same pattern as in figure 4.17

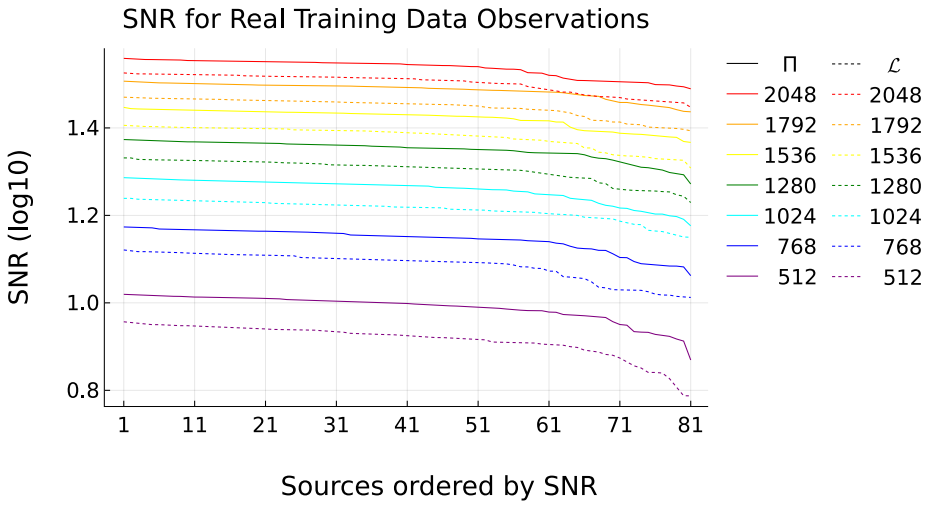

Figure 4.20: See figure 4.12 for a general description.
No strong slope can be observed at the beginning. Either the first 60 or so sources are un-mixed well, which is unlikely given the results of the other datasets, or no well un-mixed sources exists.

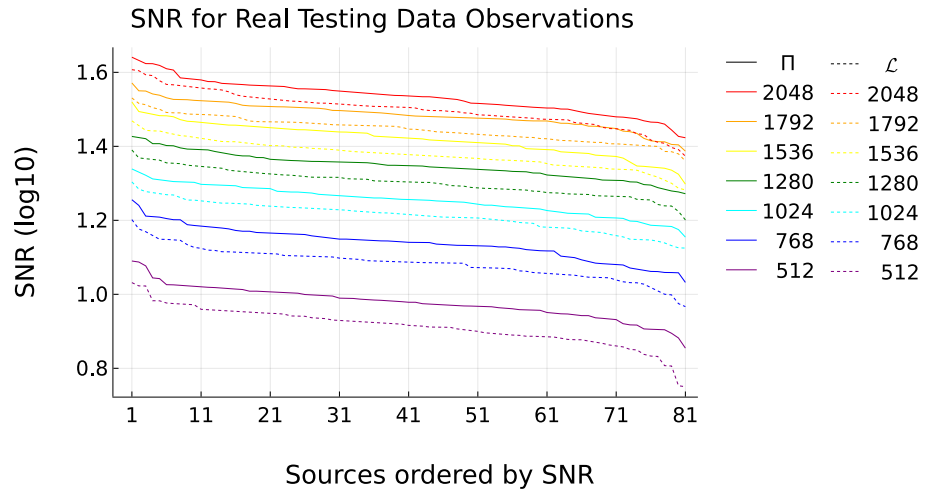

Figure 4.21: See figure 4.13 for a general description.
A few sources stand here out at the beginning in contrast to figure 4.20. Visual inspection showed that these correspond to mediocre un-mixed sources.

5 Conclusion

Summary

This thesis presented a framework which allows to run independent component analysis within a Bayesian framework with any desired prior on the mixing or un-mixing matrix. The prior designed in this thesis was concretely put on each column of the mixing matrix; however, this is an implementation detail that was employed to solve the specific inverse problem posed by EMG which was solved in this thesis. A prior on all of the mixing matrix or the un-mixing matrix directly can be used as easily within this framework.

As implied in the above, a prior for each column of the mixing matrix was derived. This was done by using a physiological EMG model to sample many possible columns of possible mixing matrices and then fitting a multivariate normal over these samples.

Furthermore, a reparametrisation of orthogonal matrices was developed which, together with pre-whitening, allowed to constrain the to-be-estimated un-mixing matrix to be always orthogonal. This was done to combat numerical instabilities due to floating point errors which arose due to the prior not being on the to-be-estimated un-mixing matrix but rather on the mixing matrix which is its inverse. The developed reparametrisation turned that problematic inverse into a much more benign transpose, thus negating the issue.

Finally, as ICA's invariance to scale and source ordering makes it difficulty to compare between results derived from different datasets or parameters or even between each estimated source, a method was presented to extract a measure from the posterior predictive distribution of the estimated MUAP traces, which allows comparison despite the drawbacks of ICA.

The above theoretical work was then verified using both synthetic and real EMG data. The results of which demonstrated the possible benefits from adding an informative prior, which was expected given similar prior work, and, of more importance for this thesis, as distinctly more novel, that analysing the Bayesian credible interval, or a measure thereof, allows to quantify which results are "bad" and which are "good".

The addition of a prior is visibly beneficial in situations where the input is very noisy or very short in terms of samples and even more so where both of these conditions coincide. As to the measure of quality, it cannot yet completely replace qualitative inspection when testing or tuning a "new" algorithm, but it certainly can be used to automate feature selection and to give a consistent measure of quality provided either "good" and "bad" outputs are distinct enough or an appropriate baseline to compare to has been determined for the experiment or application.

In conclusion, this thesis presents a method to quantify the uncertainty of the results of certain algorithms due to errors or other limitations of the input data as well as due

to inaccuracies of any real-world modelling that these algorithms may do. Furthermore, it demonstrates that this measure contains useful actionable information. Parallel to that, this thesis provides a detailed Bayesian framework together with a prior based on physiological information to run ICA on EMG data which was used to demonstrate the above.

Applicability beyond EMG

Any form of dataset, for which independent component analysis is an appropriate form of processing, should be directly usable with the framework presented in this thesis provided an appropriate prior is used. An non-informative prior is evidently always usable. If the use of an informative prior is desired, then, provided a forward model is known, that model can be used in the same way as the EMG forward model used in this thesis to estimate a prior.

Different variants of ICA or other algorithms used for source separation should also be usable provided that a maximum likelihood formulation of these exist.

Further Research

Despite measures taken to improve the computational performance, the algorithm as presented does not run fast enough to be used online, each run of the experiment took a few minutes, except maybe for small sample sizes with a low amount of channels on a very performant machine. Integration of the prior directly, if possible, into FastICA similarly to [143] may result in a significant speed-up for that part of the here presented work. Regarding the quality measures, there is less remaining room for further direct optimisation. However, it would directly benefit from improvements to the implementations of auto-differentiation and probabilistic programming used in this thesis via software libraries, see appendix B, all of which are seeing steady improvement at this time.

The model used to generate the prior for the experiments was fairly simple as the ratio of added complexity to benefit would have been unlikely to be favourable as far as the desired results where concerned. Nevertheless, as a future project, it may be worthwhile to implement a wider range of models and use the algorithms developed in this thesis to compare their performance to each other. In a similar vein, the “tanh” non-linearity used here within ICA is only one of many and an objective comparison between them may be of use.

Furthermore, the algorithms presented in chapter 3 can be applied to almost any inverse problem provided an appropriate prior can be put on $\mathbf{W}$ directly or indirectly through $\mathbf{M}$. This includes these problems arising from most other types of electrophysiological data but completely different fields of science may include problems where at the least the parts in sections 3.6, 3.7 and 3.8 can be applied.

Non-stationarity of EMG

One thing time constraints did not permit to look into was extending the algorithms of this thesis such that they can take account for non-stationarities in the signal. The non-stationarities in electromyography arise from a variety of sources: the recording electrodes are not glued or otherwise affixed to the skin and, thus, can drift along it. The skin, fat and muscle layers do move and deform independently during movement, thus, even if the recording electrodes remain fixed with regard to the skin, they would not remain stationary relative to the muscle and thus the motor units. Finally the motor unit action potentials themselves travel over time through the muscle.

The reason this movement between the locations of the motor unit action potentials and the recording electrodes is problematic follows from the fact that it is exactly their relative positions to each other within the volume conductor which determine the value of the mixing matrix and thus of the to-be-estimated un-mixing matrix. Thus, the end result of this is that the true un-mixing matrix can vary significantly over time and what is estimated, when assuming electromyographic signals to be stationary, is actually an unprincipled average over time over the "true" un-mixing matrices. The magnitude of distortion this issue causes depends heavily on the length of the data and the circumstances in which it was recorded.

This can be mitigated by using samples from shorter time windows, but, evidently, there is a direct trade-off between the amount of samples and the quality of results as can be seen well in the results presented in chapter 4.

A possibly better way would be to still use as much observation data as possible but to estimate a series of un-mixing matrices. The expected high autocorrelation between these matrices could then have been used in that case to find a way to pass information in a principled way forward and backward in time. This would have been a larger undertaking as doing so even in general is still an open research question.

Appendices

A Conference Paper

The topic of this thesis has been partially covered in a conference paper entitled
Approximate Credibility Intervals for Independent Component Analysis
by Olivier Thill and Luca Citi published in
Converging Clinical and Engineering Research on Neurorehabilitation IV, Proceedings of the 5th International Conference on Neurorehabilitation (ICNR2020), October 13–16, 2020
DOI: 10.1007/978-3-030-70316-5

B External Resources

The code for this thesis was written in the Julia programming language (julialang.org). Important software libraries used were:

- Stan (mc-stan.org)
- Flux (github.com/FluxML/Flux.jl)
- Turing (github.com/TuringLang/Turing.jl)

This of course is not an exhaustive list but covers those libraries from which ready made mathematical algorithms where used.

As stated before, the real-world datasets used were from the DeTOP repository [141].

C Symbol Glossary

This list is sorted alphabetically with Greek letters sorted according to their English names. If two meanings are given, then the symbol is used differently in separate section.

- A: Ratio of anisotropy of muscle. Here specifically the difference between the electrical conductivities parallel and orthogonal to the muscle fibres.
- $\mathbf{A}$: Matrix of coefficients describing how source signals amplitudes are mapped to electrode observations.
- α: A constant controlling the attenuation of a signal as it travels away from its origin or a variable used to scale another value.
- $\mathbf{C}$: A correlation matrix.

- c: The log of an unknown proportionality constant.
- χ: Location of an EMG electrode.
- Δ: The direction in which a variable needs to change to maximise a function.
- e: Euler's number.
- $\mathbf{e}_1$: A unit vector parallel to the first Cartesian axis.
- ϵ': The electric permittivity.
- η: A learning rate for gradient descent, It controls how much the variable to be optimised changes each iteration.
- F: The thickness of the subcutaneous fat layer.
- $\mathcal{F}$: A scalar function to be minimised or maximised.
- $\mathcal{G}$: A function to be minimised or maximised or a gamma distribution parametrised with shape $1 + \phi$ and scale $\frac{1}{\phi}$ where ϕ is the golden ratio, which has a mean and standard deviation of 1.
- $\mathbf{H}$: A matrix representing a Hessian or containing values of a similarity heuristic between two sets of vectors.
- $\mathbf{I}$: The identity matrix.
- i: Electrode Index.
- j: Source Index.
- K: A matrix rank, particularly the highest rank of multiple matrices with ascending ranks.
- k: A rank index; $1 \leq k \leq K$.
- L: Constant denoting the highest expected lag in samples.
- $\mathbf{L}$: Triangular Matrix emitted by a Cholesky decomposition.
- $\mathcal{L}$: A Likelihood function. A probability density function for all intends and purposes.
- $\mathbf{M}$: Inverse of $\mathbf{W}$. Describes a linear mixing process.
- $\mathbf{M}_0$: Inverse of $\mathbf{W}_0$.
- $\mathbf{M}_1$: Inverse and transpose of $\mathbf{W}_1$.
- $\mathcal{M}$: A model predicting the value of $\mathbf{M}$ given some parameters.

- $\mathbf{m}$: A column of $\mathbf{M}$.
- μ': The magnetic permeability.
- μ_0: The constant of vacuum permeability.
- $\mathcal{N}$: The normal distribution.
- n: An index or a number of objects.
- ω: Angular frequency of electric current.
- $\check{\mathbf{P}}$: Either a permutation or Householder matrix.
- Π: Bayesian Posterior. A probability density function.
- Π^*: Unnormalised form of Π.
- π: Bayesian Prior. A probability density function.
- $\mathbf{Q}$: An orthogonal matrix.
- $\mathbb{Q}^k$: The domain of rank k orthogonal matrices.
- $\mathbb{R}^+$: The domain of positive real numbers. Explicitly exclusive 0.
- S: The number of sources, MUs in the context of EMG.
- $\mathcal{S}$: A distribution of MU locations inside the muscle.
- $\mathbf{\Sigma}$: A covariance matrix.
- s: The thickness of the skin layer or a source index; $1 \leq s \leq S$.
- σ_{eff}: The effective electrical conductivity.
- σ_F: Electrical conductivity of fat.
- σ_S: Electrical conductivity of skin.
- σ_M: Electrical conductivity of muscle along the fibre direction.
- T: The number of samples.
- $\mathbf{T}$: An element of $\mathbb{T}$.
- $\mathbb{T}^k$: The domain of the reparametrisation of $\mathbb{Q}^k$. May be represented as rank n lower triangular matrices where each column is a unit vector or as a set of these vectors.
- Θ: The sum of parameters to be estimated.

- t: Sample index; $1 \leq t \leq T$.
- $\mathbf{t}$: An element of $\mathbf{T}$.
- υ: Location of a MUAP.
- $\mathbf{v}$: A unit vector created by rescaling a $\mathbf{w}$.
- $\mathbf{V}$: A diagonal matrix of variances.
- $\mathbf{W}$: The un-mixing matrix of dataset $\mathbf{X}$.
- $\mathbf{W}_0$: A whitening matrix of $\mathbf{X}$.
- $\mathbf{W}_1$: ICA solution of dataset $\mathbf{X}$. An orthogonal matrix.
- $\mathbf{w}$: A vector, arising from sampling such that their angles are distributed uniformly.
- X: Maximum value of x which is modelled.
- $\mathbf{X}$: A matrix with s rows, each being an observation over time, and t columns, each being a moment in time.
- $\mathbf{X}_0$: A whitened $\mathbf{X}$; $\mathbf{X}_0 = \mathbf{W}_0\mathbf{X}$.
- Ξ: Data which carries information about Θ. This refers to the concept, concrete instances are referred to as the matrix $\mathbf{X}$.
- x: The distance along an axis parallel to the skin surface orthogonal to the axis of y.
- Y: Maximum value of y which is modelled.
- $\mathbf{Y}$: A matrix with s rows, each being an estimated source over time, and t columns, each being a moment in time.
- y: The distance along an axis parallel to the skin surface orthogonal to the axis of x.
- Z: Maximum value of z which is modelled.
- z: The distance along an axis orthogonal to the skin surface pointing downwards starting from the surface of the muscle.
- z_t: Total depth; $z_t = z + S + F$.

D Bibliography

[1] Mads Jochumsen et al. "Movement intention detection in adolescents with cerebral palsy from single-trial EEG". In: *Journal of Neural Engineering* 15.6 (Oct. 2018), p. 066030. DOI: 10.1088/1741-2552/aae4b8.

[2] Surjo R. Soekadar, Klaus Haagen, and Niels Birbaumer. "Brain-Computer Interfaces (BCI): Restoration of Movement and Thought from Neuroelectric and Metabolic Brain Activity". In: *Coordination: Neural, Behavioral and Social Dynamics*. Ed. by Armin Fuchs and Viktor K. Jirsa. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 229–252. ISBN: 978-3-540-74479-5. DOI: 10.1007/978-3-540-74479-5_11.

[3] Zhijun Li et al. "Hybrid Brain/Muscle Signals Powered Wearable Walking Exoskeleton Enhancing Motor Ability in Climbing Stairs Activity". In: *IEEE Transactions on Medical Robotics and Bionics* 1.4 (Nov. 2019), pp. 218–227. DOI: 10.1109/TMRB.2019.2949865.

[4] Minkyu Ahn et al. "A Review of Brain-Computer Interface Games and an Opinion Survey from Researchers, Developers and Users". In: *Sensors* 14.8 (Aug. 2014), pp. 14601–14633. DOI: 10.3390/s140814601.

[5] Daniel C. Miller and Philip A. Defina. "The Application of Neuroscience to the Practice of School Neuropsychology". In: *Best Practices in School Neuropsychology*. John Wiley & Sons, Oct. 19, 2009.

[6] Juri D. Kropotov. *Quantitative EEG, event-related potentials and neurotherapy*. Amsterdam Boston London: Elsevier/Academic, 2009. ISBN: 9780080922973.

[7] Chee Siang Ang et al. "Use of brain computer interfaces in neurological rehabilitation". In: *British Journal of Neuroscience Nursing* 7.3 (June 2011), pp. 523–528. DOI: 10.12968/bjnn.2011.7.3.523.

[8] D. Farina et al. "The Extraction of Neural Information from the Surface EMG for the Control of Upper-Limb Prostheses: Emerging Avenues and Challenges". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.4 (July 2014), pp. 797–809. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2014.2305111.

[9] Ethan Buch et al. "Think to Move: a Neuromagnetic Brain-Computer Interface (BCI) System for Chronic Stroke". In: *Stroke* 39.3 (Mar. 2008), pp. 910–917. DOI: 10.1161/strokeaha.107.505313.

[10] Mingxing Lyu et al. "Development of an EMG-Controlled Knee Exoskeleton to Assist Home Rehabilitation in a Game Context". In: *Frontiers in Neurorobotics* 13 (Aug. 2019). DOI: 10.3389/fnbot.2019.00067.

[11] Claudio Castellini. "Upper Limb Active Prosthetic systems—Overview". In: *Wearable Robotics*. Elsevier, 2020, pp. 365–376. DOI: 10.1016/b978-0-12-814659-0.00019-9.

[12] S. Raspopovic et al. "Restoring Natural Sensory Feedback in Real-Time Bidirectional Hand Prostheses". In: *Science Translational Medicine* 6.222 (Feb. 2014), 222ra19–222ra19. DOI: 10.1126/scitranslmed.3006820.

[13] Morten B. Kristoffersen et al. "The Effect of Feedback During Training Sessions on Learning Pattern-Recognition-Based Prosthesis Control". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.10 (Oct. 2019), pp. 2087–2096. DOI: 10.1109/tnsre.2019.2929917.

[14] Wm. H. Dobelle. "Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex". In: *ASAIO Journal: Artificial Organ Research and Development* 46.1 (2000). ISSN: 1058-2916.

[15] Aya Rezeika et al. "Brain–Computer Interface Spellers: A Review". In: *Brain Sciences* 8.4 (Mar. 2018), p. 57. DOI: 10.3390/brainsci8040057.

[16] L. Citi et al. "P300-Based BCI Mouse With Genetically-Optimized Analogue Control". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16.1 (Feb. 2008), pp. 51–61. DOI: 10.1109/tnsre.2007.913184.

[17] Tomislav Milekovic et al. "Stable long-term BCI-enabled communication in ALS and locked-in syndrome using LFP signals". In: *Journal of Neurophysiology* 120.1 (July 2018), pp. 343–360. DOI: 10.1152/jn.00493.2017.

[18] Martin Spüler. "Questioning the evidence for BCI-based communication in the complete locked-in state". In: *PLOS Biology* 17.4 (Apr. 2019). Ed. by Ulrich Dirnagl, e2004750. DOI: 10.1371/journal.pbio.2004750.

[19] Robert Bogue. "Exoskeletons and robotic prosthetics: a review of recent developments". In: *Industrial Robot: An International Journal* 36.5 (Aug. 2009), pp. 421–427. DOI: 10.1108/01439910910980141.

[20] Caterina Cinel, Davide Valeriani, and Riccardo Poli. "Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects". In: *Frontiers in Human Neuroscience* 13 (Jan. 2019). DOI: 10.3389/fnhum.2019.00013.

[21] Riccardo Poli, Davide Valeriani, and Caterina Cinel. "Collaborative Brain-Computer Interface for Aiding Decision-Making". In: *PLoS ONE* 9.7 (July 2014). Ed. by Maurice J. Chacron, e102693. DOI: 10.1371/journal.pone.0102693.

[22] Gabriel Alves Mendes Vasiljevic and Leonardo Cunha de Miranda. "Brain–Computer Interface Games Based on Consumer-Grade EEG Devices: A Systematic Literature Review". In: *International Journal of Human–Computer Interaction* 36.2 (June 2019), pp. 105–142. DOI: 10.1080/10447318.2019.1612213.

[23] Raffaella Folgieri, Mattia G. Bergomi, and Simone Castellani. "EEG-Based Brain-Computer Interface for Emotional Involvement in Games Through Music". In: *Digital Da Vinci*. Springer New York, 2014, pp. 205–236. DOI: 10.1007/978-1-4939-0536-2_9.

[24] Ranit Sengupta et al. "Theta Modulated Neural Phase Coherence Facilitates Speech Fluency in Adults Who Stutter". In: *Frontiers in Human Neuroscience* 13 (Nov. 2019). DOI: 10.3389/fnhum.2019.00394.

[25] Abdoreza Asadpour, Mehran Jahed, and Saeid Mahmoudian. "Brain Waves Evaluation of Sound Therapy in Chronic Subjective Tinnitus Cases Using Wavelet Decomposition". In: *Frontiers in Integrative Neuroscience* 12 (Sept. 2018). DOI: 10.3389/fnint.2018.00038.

[26] Fleur Margaret Howells, Dan J. Stein, and Vivienne A. Russell. "Childhood Trauma is Associated with Altered Cortical Arousal: Insights from an EEG Study". In: *Frontiers in Integrative Neuroscience* 6 (2012). DOI: 10.3389/fnint.2012.00120.

[27] Dae-Keun Kim, Jyoo-Hi Rhee, and Seung Wan Kang. "Reorganization of the brain and heart rhythm during autogenic meditation". In: *Frontiers in Integrative Neuroscience* 7 (2014). DOI: 10.3389/fnint.2013.00109.

[28] Thalía Harmony. "The functional significance of delta oscillations in cognitive processing". In: *Frontiers in Integrative Neuroscience* 7 (2013). DOI: 10.3389/fnint.2013.00083.

[29] Max Heinrich Fischer and Hans Löwenbach. "Aktionsströme des Zentralnervensystems unter der Einwirkung von Krampfgiften". In: *Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie* 174.3-4 (1933), pp. 357–382. DOI: 10.1007/bf01864468.

[30] Jeffrey R. Cram and Jeffrey C. Steger. "EMG scanning in the diagnosis of chronic pain". In: *Biofeedback and Self-Regulation* 8.2 (June 1983), pp. 229–241. DOI: 10.1007/bf00998853.

[31] M. Brandon Westover et al. "Diagnostic Value of Electroencephalography with Ten Electrodes in Critically Ill Patients". In: *Neurocritical Care* (Feb. 2020). DOI: 10.1007/s12028-019-00911-4.

[32] Francesco Pisani, Carlotta Spagnoli, and Carlo Fusco. "EEG Monitoring of the Epileptic Newborn". In: *Current Neurology and Neuroscience Reports* 20.4 (Mar. 2020). DOI: 10.1007/s11910-020-1027-7.

[33] C. S. Pattichis, C. N. Schizas, and L. T. Middleton. "Neural network models in EMG diagnosis". In: *IEEE Transactions on Biomedical Engineering* 42.5 (May 1995), pp. 486–496. DOI: 10.1109/10.376153.

[34] Abdulhamit Subasi. "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders". In: *Computers in Biology and Medicine* 43.5 (June 2013), pp. 576–586. DOI: 10.1016/j.compbiomed.2013.01.020.

[35] Hiroyuki Kidokoro et al. "High-amplitude fast activity in EEG: An early diagnostic marker in children with beta-propeller protein-associated neurodegeneration (BPAN)". In: *Clinical Neurophysiology* (June 2020). DOI: 10.1016/j.clinph.2020.06.006.

[36] Jessica Joanne van der Zande et al. "Diagnostic and prognostic value of EEG in prodromal dementia with Lewy bodies". In: *Neurology* (July 2020), 10.1212/WNL.0000000000009977. DOI: 10.1212/wnl.0000000000009977.

[37] Álvaro Gutiérrez et al. "Serious Game Platform with Haptic Feedback and EMG Monitoring for Upper Limb Rehabilitation and Smoothness Quantification on Spinal Cord Injury Patients". In: *Applied Sciences* 10.3 (Feb. 2020), p. 963. DOI: 10.3390/app10030963.

[38] Rui Sun, Rong Song, and Kai-yu Tong. "Complexity Analysis of EMG Signals for Patients After Stroke During Robot-Aided Rehabilitation Training Using Fuzzy Approximate Entropy". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.5 (Sept. 2014), pp. 1013–1019. DOI: 10.1109/tnsre.2013.2290017.

[39] Francisco Sepulveda. "Brain-actuated Control of Robot Navigation". In: *Advances in Robot Navigation*. Ed. by Alejandra Barrera. July 5, 2011. ISBN: 978-953-307-346-0. DOI: 10.5772/17401.

[40] M. van Gerven et al. "The brain-computer interface cycle". In: *Journal of Neural Engineering* 6.4 (Aug. 2009), p. 041001. DOI: 10.1088/1741-2560/6/4/041001.

[41] D. J. McFarland et al. "BCI Meeting 2005–workshop on BCI signal processing: feature extraction and translation". In: *IEEE Trans Neural Syst Rehabil Eng* 14.2 (June 2006), pp. 135–138. DOI: 10.1109/tnsre.2006.875637.

[42] R. Menon et al. "Study on interaction between temporal and spatial information in classification of EMG signals in myoelectric prostheses". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP.99 (2017), pp. 1–1. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2017.2687761.

[43] Asrul Adam et al. "Feature Selection and Classifier Parameters Estimation for EEG Signals Peak Detection Using Particle Swarm Optimization". In: *The Scientific World Journal* 2014 (2014), pp. 1–13. DOI: 10.1155/2014/973063.

[44] A. Bashashati et al. "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals". In: *Journal of Neural Engineering* 4.2 (June 2007), pp. 32–57. DOI: 10.1088/1741-2560/4/2/r03.

[45] K. S. Hong and H. Santosa. "Current BCI technologies in brain engineering". In: *2013 International Conference on Robotics, Biomimetics, Intelligent Computational Systems*. Nov. 2013, pp. 1–4. DOI: 10.1109/ROBIONETICS.2013.6743567.

[46] O. A. Padierna Sosa et al. "BCI: A historical analysis and technology comparison". In: *2011 Pan American Health Care Exchanges*. IEEE, Mar. 2011. DOI: 10.1109/pahce.2011.5871883.

[47] Fernando Lopes Silva. "EEG: Origin and Measurement". In: *EEG - fMRI Physiological Basis, Technique, and Applications*. Ed. by Christoph Mulert and Louis Lemieux. Springer-Verlag Berlin Heidelberg, Oct. 29, 2009, pp. 19–38. ISBN: 978-3-540-87919-0. DOI: 10.1007/978-3-540-87919-0.

[48] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin. "Techniques of EMG signal analysis: detection, processing, classification and applications". In: *Biological Procedures Online* 8.1 (2006), pp. 11–35. ISSN: 1480-9222. DOI: 10.1251/bpo115.

[49] Harvey Lodish et al. *Molecular Cell Biology. 4th edition.* New York: W.H. Freeman, 2000. ISBN: 0716731363.

[50] Patrick Ledwidge, Jeremy Foust, and Adam Ramsey. "Recommendations for Developing an EEG Laboratory at a Primarily Undergraduate Institution." In: *Journal of undergraduate neuroscience education : JUNE : a publication of FUN, Faculty for Undergraduate Neuroscience* 17 (1 2018), A10–A19. ISSN: 1544-2896. epublish.

[51] Alice F. Jackson and Donald J. Bolger. "The neurophysiological bases of EEG and EEG measurement: A review for the rest of us". In: *Psychophysiology* 51.11 (July 2014), pp. 1061–1071. DOI: 10.1111/psyp.12283.

[52] Robert Oostenveld and Peter Praamstra. "The five percent electrode system for high-resolution EEG and ERP measurements". In: *Clinical Neurophysiology* 112.4 (Apr. 2001), pp. 713–719. DOI: 10.1016/s1388-2457(00)00527-7.

[53] Bin He, Yunhua Wang, and Dongsheng Wu. "Estimating cortical potentials from scalp EEGs in a realistically shaped inhomogeneous head model by means of the boundary element method". In: *IEEE Transactions on Biomedical Engineering* 46.10 (Oct. 1999), pp. 1264–1268. ISSN: 0018-9294. DOI: 10.1109/10.790505.

[54] C. Nordander et al. "Influence of the subcutaneous fat layer, as measured by ultrasound, skinfold calipers and BMI, on the EMG amplitude". In: *European Journal of Applied Physiology* 89.6 (2003), pp. 514–519. DOI: 10.1007/s00421-003-0819-1.

[55] A. Teklemariam et al. “A Finite Element Model simulation of surface EMG signals based on muscle tissue dielectric properties and electrodes configuration”. In: *2014 9thInternational Symposium on Communication Systems, Networks Digital Sign (CSNDSP)*. July 2014, pp. 193–197. DOI: 10.1109/CSNDSP.2014.6923823.

[56] Dennis J. McFarland et al. “Spatial filter selection for EEG-based communication”. In: *Electroencephalography and Clinical Neurophysiology* 103.3 (Sept. 1997), pp. 386–394. DOI: 10.1016/s0013-4694(97)00022-2.

[57] SanjayP Singh. “Magnetoencephalography: Basic principles”. In: *Annals of Indian Academy of Neurology* 17.5 (2014), p. 107. DOI: 10.4103/0972-2327.128676.

[58] A Villringer. “Non-invasive optical spectroscopy and imaging of human brain function”. In: *Trends in Neurosciences* 20.10 (Oct. 1997), pp. 435–442. DOI: 10.1016/s0166-2236(97)01132-6.

[59] L. R. Hochberg and J. P. Donoghue. “Sensors for brain-computer interfaces”. In: *IEEE Engineering in Medicine and Biology Magazine* 25.5 (Sept. 2006), pp. 32–38. ISSN: 0739-5175. DOI: 10.1109/MEMB.2006.1705745.

[60] Johannes Kögel et al. “Using brain-computer interfaces: a scoping review of studies employing social research methods”. In: *BMC Medical Ethics* 20.1 (Mar. 2019). DOI: 10.1186/s12910-019-0354-1.

[61] Eran Klein. “Informed Consent in Implantable BCI Research: Identifying Risks and Exploring Meaning”. In: *Science and Engineering Ethics* 22.5 (Oct. 2015), pp. 1299–1317. DOI: 10.1007/s11948-015-9712-7.

[62] Timothy Brown et al. “Personal responsibility in the age of user-controlled neuroprosthetics”. In: *2016 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*. IEEE, May 2016. DOI: 10.1109/ethics.2016.7560039.

[63] QianQian Li, Ding Ding, and Mauro Conti. “Brain-Computer Interface applications: Security and privacy challenges”. In: *2015 IEEE Conference on Communications and Network Security (CNS)*. IEEE, Sept. 2015. DOI: 10.1109/cns.2015.7346884.

[64] Ivan Martinovic et al. “On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces”. In: *21st USENIX Security Symposium (USENIX Security 12)*. Bellevue, WA: USENIX Association, Aug. 2012, pp. 143–158. ISBN: 978-1-931971-95-9.

[65] Roberto Merletti and Dario Farina, eds. *Surface Electromyography : Physiology, Engineering, and Applications*. John Wiley & Sons, Inc., Apr. 2016. 592 pp. ISBN: 1118987020. DOI: 10.1002/9781119082934.

[66] Roberto Merletti and Philip Parker, eds. *Electromyography*. John Wiley & Sons, Inc., July 2004. DOI: 10.1002/0471678384.

[67] E. Chauvet, O. Fokapu, and D. Gamet. “Inverse problem in the surface EMG: a feasibility study”. In: *2001 Conference Proceedings of the 23rdAnnual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 2. 2001, pp. 1048–1050. DOI: 10.1109/IEMBS.2001.1020368.

[68] B. S. Darak and S. M. Hambarde. “A review of techniques for extraction of cardiac artifacts in surface EMG signals and results for simulation of ECG-EMG mixture signal”. In: *2015 International Conference on Pervasive Computing (ICPC)*. Jan. 2015, pp. 1–5. DOI: 10.1109/PERVASIVE.2015.7087067.

[69] N. S. Stoykov et al. “Frequency- and time-domain FEM models of EMG: capacitive effects and aspects of dispersion”. In: *IEEE Transactions on Biomedical Engineering* 49.8 (Aug. 2002), pp. 763–772. DOI: 10.1109/tbme.2002.800754.

[70] H. P. Schwan. “Electrical properties of tissues and cell suspensions: mechanisms and models. a”. In: *Proceedings of 16thAnnual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 1. Nov. 1994, pp. 70–71. DOI: 10.1109/IEMBS.1994.412155.

[71] Steven Ovadia and Morad Askari. “Upper Extremity Amputations and Prosthetics”. In: *Seminars in Plastic Surgery* 29.01 (Feb. 2015), pp. 055–061. DOI: 10.1055/s-0035-1544171.

[72] Cynthia Furse, Douglas A. Christensen, and Carl H. Durney. *Basic Introduction to Bioelectromagnetics, Second Edition*. CRC Press, 2009. ISBN: 978-1420055429.

[73] Bert A. Albers et al. “A Model Study on the Influence of Structure and Membrane Capacitance on Volume Conduction in Skeletal Muscle Tissue”. In: *IEEE Transactions on Biomedical Engineering* BME-33.7 (July 1986), pp. 681–689. DOI: 10.1109/tbme.1986.325758.

[74] C. Gabriel and S. Gabriel. *Compilation of the Dielectric Properties of Body Tissues at RF and Microwave Frequencies*. Dept of Physics - King’s College London, 1995. DOI: 10.21236/ada303903.

[75] L. A. Geddes and L. E. Baker. “The specific resistance of biological material–a compendium of data for the biomedical engineer and physiologist”. In: *Medical and Biological Engineering* 5.3 (May 1967), pp. 271–293. DOI: 10.1007/bf02474537.

[76] F. L. Gielen, W. Wallinga-de Jonge, and K. L. Boon. “Electrical conductivity of skeletal muscle tissue: experimental results from different muscles in vivo”. In: *Medical and Biological Engineering and Computing* 22.6 (Nov. 1984), pp. 569–577. DOI: 10.1007/bf02443872.

[77] Damijan Miklavcic, Natasa Pavselj, and Francis X. Hart. “Electric Properties of Tissues”. In: *Wiley Encyclopedia of Biomedical Engineering*. Vol. 6. Apr. 2006. ISBN: 9780471740360. DOI: 10.1002/9780471740360.ebs0403.

[78] Ilja L. Kruglikov. “Influence of layered skin structure on the distribution of radiofrequency currents in dermis and subcutaneous fat”. In: *AIP Advances* 5.12 (Dec. 2015), p. 127122. DOI: 10.1063/1.4938545.

[79] M. M. Lowery et al. “A multiple-layer finite-element model of the surface EMG signal”. In: *IEEE Transactions on Biomedical Engineering* 49.5 (2002), pp. 446–454. DOI: 10.1109/10.995683.

[80] B. Feinstein et al. “Morphologic studies of motor units in normal human muscles”. In: *Acta Anatomica* 23.2 (1955), pp. 127–142. DOI: 10.1159/000140989.

[81] Laura Miller McPherson et al. “Properties of the motor unit action potential shape in proximal and distal muscles of the upper limb in healthy and post-stroke individuals”. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Aug. 2016. DOI: 10.1109/embc.2016.7590708.

[82] S. Methenitis et al. “Muscle Fiber Conduction Velocity, Muscle Fiber Composition, and Power Performance”. In: *Medicine and Science in Sports and Exercise* 48.9 (Sept. 2016), pp. 1761–1771. DOI: 10.1249/mss.0000000000000954.

[83] Steen Andreassen and Lars Arendt-Nielsen. "Muscle fibre conduction velocity in motor units of the human anterior tibial muscle: a new size principle parameter". In: *The Journal of Physiology* 391 (Oct. 1987), pp. 561–571. DOI: 10.1113/jphysiol.1987.sp016756.

[84] Ignacio Rodríguez Carreño, Luis Gila-Useros, and Armando Malanda-Trigueros. "Motor Unit Action Potential Duration: Measurement and Significance". In: *Advances in Clinical Neurophysiology*. Ed. by Mohammad Abud Ajeena Dr. Ihsan. InTech, 2012. DOI: 10.5772/50265.

[85] A. Gechev et al. "Potential risks of iatrogenic complications of nerve conduction studies (NCS) and electromyography (EMG)". In: *Clinical Neurophysiology Practice* 1.Supplement C (2016), pp. 62–66. ISSN: 2467-981X. DOI: 10.1016/j.cnp.2016.09.003.

[86] Roberto Merletti and Dario Farina. "Analysis of intramuscular electromyogram signals". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367.1887 (2009), pp. 357–368. ISSN: 1364-503X. DOI: 10.1098/rsta.2008.0235. eprint: http://rsta.royalsocietypublishing.org/content/367/1887/357.full.pdf.

[87] Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. "Fractal analysis features for weak and single-channel upper-limb EMG signals". In: *Expert Systems with Applications* 39.12 (2012), pp. 11156–11163. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2012.03.039.

[88] Roberto Merletti, Ales Holobar, and Dario Farina. "Two dimensional high density surface EMG (HD-EMG) technology and applications". In: *Proceedings, XVIIth Congress of the International Society of Electrophysiology and Kinesiology (ISEK), 18-21 June 2008, Niagara Falls, Canada [CD-ROM]*. ISEK, 2008.

[89] A. Holobar, M. A. Minetto, and D. Farina. "Accurate identification of motor unit discharge patterns from high-density surface EMG and validation with a novel signal-based performance metric". In: *Journal of Neural Engineering* 11.1 (Jan. 2014), p. 016008. DOI: 10.1088/1741-2560/11/1/016008.

[90] Andrzej Wolski. "Theory of electromagnetic fields". In: (Nov. 18, 2011). arXiv: 1111.4354.

[91] R. Merletti et al. "Modeling of surface myoelectric signals. I. Model implementation". In: *IEEE Transactions on Biomedical Engineering* 46.7 (1999), pp. 810–820. DOI: 10.1109/10.771190.

[92] A. Heringa, D. F. Stegeman, and J. P. C. de Weerd. "Calculated potential and electric field distributions around an active nerve fiber". In: *Journal of Applied Physics* 66.6 (1989), pp. 2724–2731. DOI: 10.1063/1.344213.

[93] T. H. J. M. Gootzen, D. F. Stegeman, and A. Van Oosterom. "Finite limb dimensions and finite muscle length in a model for the generation of electromyographic signals". In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 81.2 (1991), pp. 152–162. DOI: 10.1016/0168-5597(91)90008-l.

[94] J. H. Blok, D. F. Stegeman, and A. van Oosterom. "Three-Layer Volume Conductor Model and Software Package for Applications in Surface Electromyography". In: *Annals of Biomedical Engineering* 30.4 (Apr. 1, 2002), pp. 566–577. ISSN: 1573-9686. DOI: 10.1114/1.1475345.

[95] Eike Petersen and Philipp Rostalski. "A Comprehensive Mathematical Model of Motor Unit Pool Organization, Surface Electromyography, and Force Generation". In: *Frontiers in Physiology* 10 (Mar. 2019). DOI: 10.3389/fphys.2019.00176.

[96] L. Mesin. "Simulation of surface EMG signals for a multilayer volume conductor with a superficial bone or blood vessel". In: *IEEE Transactions on Biomedical Engineering* 55.6 (June 2008), pp. 1647–1657. DOI: 10.1109/tbme.2008.919104.

[97] Babuska Szabo. *Finite Element Analysis*. John Wiley & Sons, Inc., Mar. 29, 1991. 388 pp. ISBN: 0471502731.

[98] J. P. van Dijk et al. "Evidence of Potential Averaging over the Finite Surface of a Bioelectric Surface Electrode". In: *Annals of Biomedical Engineering* 37.6 (2009), pp. 1141–1151. DOI: 10.1007/s10439-009-9680-7.

[99] Pierre Comon and Christian Jutten, eds. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Pr Inc, Apr. 21, 2010. 831 pp. ISBN: 978-0-12-374726-6.

[100] Christopher J. James and Christian W. Hesse. "Independent component analysis for biomedical signals". In: *Physiological Measurement* 26.1 (Dec. 2004), R15–R39. DOI: 10.1088/0967-3334/26/1/r02.

[101] Edwin F. Taylor. "Why does nothing move faster than light? Because ahead is ahead!" In: *American Journal of Physics* 58.9 (Sept. 1990), pp. 889–890. DOI: 10.1119/1.16343.

[102] C. Lamus et al. "An Analysis of How Spatiotemporal Dynamic Models of Brain Activity Could Improve MEG/EEG Inverse Solutions". In: *ArXiv e-prints* (Nov. 2015). arXiv: 1511.03726 [stat.AP].

[103] X. Zhu and Y. Zhang. "High-density surface EMG decomposition based on a convolutive blind source separation approach". In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Aug. 2012, pp. 609–612. DOI: 10.1109/EMBC.2012.6346005.

[104] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. "Independent Component Analysis". In: *Independent Component Analysis*. .9. John Wiley & Sons, Inc., May 21, 2001. Chap. ICA by Maximum Likelihood Estimation, pp. 203–219. DOI: 10.1002/0471221317.

[105] S. H. Nawab, S. S. Chang, and C. J. De Luca. "High-yield decomposition of surface EMG signals". In: *Clinical Neurophysiology* 121.10 (Oct. 2010), pp. 1602–1615. DOI: 10.1016/j.clinph.2009.11.092.

[106] Karl Pearson. "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (Nov. 1901), pp. 559–572. DOI: 10.1080/14786440109462720.

[107] Agnan Kessy, Alex Lewin, and Korbinian Strimmer. "Optimal Whitening and Decorrelation". In: *The American Statistician* 72.4 (Jan. 2018), pp. 309–314. DOI: 10.1080/00031305.2016.1277159.

[108] Mike E. Davies et al., eds. *Independent Component Analysis and Signal Separation*. Springer, Dec. 20, 2007. 872 pp. ISBN: 978-3-540-74494-8. DOI: 10.1007/978-3-540-74494-8.

[109] K. Friston et al. “Multiple sparse priors for the M/EEG inverse problem”. In: *Neuroimage* 39.3 (Feb. 2008), pp. 1104–1120. DOI: 10.1016/j.neuroimage.2007.09.048.

[110] M. Plumbley. “Conditions for nonnegative independent component analysis”. In: *IEEE Signal Processing Letters* 9.6 (June 2002), pp. 177–180. ISSN: 1070-9908. DOI: 10.1109/LSP.2002.800502.

[111] Jean-François Cardoso. “Dependence, correlation and gaussianity in independent component analysis”. In: *Journal of Machine Learning Research* 4.Dec (Dec. 2013), pp. 1177–1203.

[112] A. Hyvarinen. “Fast and robust fixed-point algorithms for independent component analysis”. In: *IEEE Transactions on Neural Networks* 10.3 (May 1999), pp. 626–634. DOI: 10.1109/72.761722.

[113] Jari Miettinen et al. “The squared symmetric FastICA estimator”. In: *Signal Processing 131 (2017) 402-411* (Dec. 17, 2015). DOI: 10.1016/j.sigpro.2016.08.028. arXiv: 1512.05534 [math.ST].

[114] K. Nordhausen et al. “Deflation-based FastICA reloaded”. In: *2011 19th European Signal Processing Conference*. Aug. 2011, pp. 1854–1858.

[115] Jari Miettinen et al. “Deflation-Based FastICA With Adaptive Choices of Nonlinearities”. In: *IEEE Transactions on Signal Processing* 62.21 (Nov. 2014), pp. 5716–5724. DOI: 10.1109/tsp.2014.2356442.

[116] W. Wu, S. Nagarajan, and Z. Chen. “Bayesian Machine Learning: EEG/MEG signal processing measurements”. In: *IEEE Signal Processing Magazine* 33.1 (Jan. 2016), pp. 14–36. ISSN: 1053-5888. DOI: 10.1109/msp.2015.2481559.

[117] C. W. Hesse and C. J. James. “On Semi-Blind Source Separation Using Spatial Constraints With Applications in EEG Analysis”. In: *IEEE Transactions on Biomedical Engineering* 53.12 (Dec. 2006), pp. 2525–2534. DOI: 10.1109/tbme.2006.883796.

[118] Ganesh R. Naik et al. “Hand gestures for HCI using ICA of EMG”. In: *VisHCI 06 Proceedings of the HCSNet workshop on Use of vision in human-computer interaction - Volume 56*. Ed. by editor. 2006, pp. 67–72.

[119] Ganesh R. Naik et al. “Subtle Hand Gesture Identification for HCI Using Temporal Decorrelation Source Separation BSS of Surface EMG”. In: *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*. IEEE, Dec. 2007. DOI: 10.1109/dicta.2007.4426772.

[120] G. R. Naik, D. K. Kumar, and M. Palaniswami. “Addressing source separation and identification issues in surface EMG using blind source separation”. In: *Conf Proc IEEE Eng Med Biol Soc* 2008 (2008), pp. 1124–1127. DOI: 10.1109/iembs.2008.4649358.

[121] Ganesh R. Naik, Dinesh K. Kumar, and Marimuthu Palaniswami. “Multi run ICA and surface EMG based signal processing system for recognising hand gestures”. In: *2008 8th IEEE International Conference on Computer and Information Technology*. IEEE, July 2008. DOI: 10.1109/cit.2008.4594760.

[122] Dario Farina and Alberto Rainoldi. “Compensation of the effect of sub-cutaneous tissue layers on surface EMG: a simulation study”. In: *Medical Engineering & Physics* 21.6-7 (1999), pp. 487–497. ISSN: 1350-4533. DOI: 10.1016/s1350-4533(99)00075-2.

[123] Stephen Roberts and Rizwan Choudrey. "Bayesian Independent Component Analysis with Prior Constraints: An Application in Biosignal Analysis". In: *Lecture Notes in Computer Science*. Ed. by Joab Winkler, Mahesan Niranjan, and Neil Lawrence. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 159–179. ISBN: 978-3-540-31728-9. DOI: 10.1007/11559887_10.

[124] Pierre-Simon de Laplace. *Oeuvres complètes de Laplace [Texte imprimé] / publiées sous les auspices de l'Académie des sciences, par MM. les secrétaires perpétuels*. Ed. by Gauthier-Villars. 1782.

[125] Andrew Gelman et al. *Bayesian Data Analysis*. Chapman and Hall/CRC, June 1995. DOI: 10.1201/9780429258411.

[126] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: (Sept. 15, 2016). arXiv: 1609.04747 [cs.LG].

[127] Michalis K. Titsias, Magnus Rattray, and Neil D. Lawrence. "Markov chain Monte Carlo algorithms for Gaussian processes". In: *Bayesian Time Series Models*. Ed. by David Barber, A. Taylan Cemgil, and SilviaEditors Chiappa. Cambridge University Press, 2011, pp. 295–316. DOI: 10.1017/CBO9780511984679.015.

[128] E. B. Dynkin. *Theory of Markov Processes*. Elsevier Science, May 12, 2014. 220 pp. ISBN: 9781483226101.

[129] D. Hofmann et al. "Bayesian Filtering of Surface EMG for Accurate Simultaneous and Proportional Prosthetic Control". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24.12 (Dec. 2016), pp. 1333–1341. DOI: 10.1109/tnsre.2015.2501979.

[130] Dirk P. Kroese et al. "Why the Monte Carlo method is so important today". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 6.6 (June 2014), pp. 386–392. DOI: 10.1002/wics.1314.

[131] S. Brooks et al. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2011. ISBN: 9781420079425.

[132] V. I. Arnol'd. *Mathematical Methods of Classical Mechanics*. Springer New York, Sept. 5, 1997. 540 pp. ISBN: 0387968903.

[133] Michael Betancourt. "A Conceptual Introduction to Hamiltonian Monte Carlo". In: (Jan. 10, 2017). arXiv: 1701.02434 [stat.ME].

[134] Ludger Starke and Dirk Ostwald. "Variational Bayesian Parameter Estimation Techniques for the General Linear Model". In: *Frontiers in Neuroscience* 11 (Sept. 2017). DOI: 10.3389/fnins.2017.00504.

[135] S. Geisser et al. "The validity of posterior expansions based on Laplace's method". In: *Bayesian and likelihood methods in statistics and econometrics* 7 (1990), p. 473.

[136] Donald Knuth. *Stable marriage and its relation to other combinatorial problems : an introduction to the mathematical analysis of algorithms*. Providence, R.I: American Mathematical Society, 1997. ISBN: 9780821806036.

[137] D. Gale and L. S. Shapley. "College Admissions and the Stability of Marriage". In: *The American Mathematical Monthly* 69.1 (Jan. 1962), p. 9. DOI: 10.2307/2312726.

[138] Timothy Dozat. "Incorporating nesterov momentum into adam". In: (2016).

[139] Gene Golub. *Matrix computations*. Baltimore: The Johns Hopkins University Press, 2013. ISBN: 9781421408590.

[140] Larry Grove. *Classical groups and geometric algebra*. Providence, Rhode Island: American Mathematical Society, 2002. ISBN: 9781470420918.

[141] Ana Matran-Fernandez et al. "SEEDS dataset". In: (2020). DOI: 10.17605/OSF.IO/WA3QK.

[142] Ana Matran-Fernandez et al. "SEEDS, simultaneous recordings of high-density EMG and finger joint angles during multiple hand movements". In: *Scientific Data* 6.1 (Sept. 2019), p. 186. ISSN: 2052-4463. DOI: 10.1038/s41597-019-0200-9.

[143] C. W. Hesse and C. J. James. "The FastICA algorithm with spatial constraints". In: *IEEE Signal Processing Letters* 12.11 (Nov. 2005), pp. 792–795. DOI: 10.1109/lsp.2005.856867.