Dynamic Magnetic Resonance Elastography:
Improved Direct Methods of Shear Modulus Estimation

by

C. Antonio Sánchez

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Magnetic Resonance Elastography (MRE) is a medical imaging technique used to generate a map of tissue elasticity. The resulting image is known as an elastogram, and gives a quantitative measure of stiffness in the examined tissue. The method is indirect; the elasticity, itself, is not measured. Instead, the physical response to a known stress is captured using magnetic resonance imaging, and is related to an elasticity parameter through a mathematical model of the tissue. In dynamic elastography, a harmonic stress is externally applied by a mechanical actuator, which is oriented to induce shear waves through the tissue. Once the system reaches a quasi-steady state, the displacement field is measured at a sequence of points in time. This data is the input to elasticity reconstruction algorithms. In this dissertation, the tissue is modelled as a linearly viscoelastic, isotropic continuum, undergoing harmonic motion with a known fundamental frequency. With this model, viscoelasticity is described by the complex versions of Lamé's first and second parameters. The second parameter, known as the complex shear modulus, is the one of interest. The term involving the first parameter is usually deemed negligible, so is ignored.

The task is to invert the tissue model, a system of linear differential equations, to find the desired parameter. Direct inversion methods use the measured data directly in the model. Most current direct methods assume the shear modulus can be approximated locally by a constant, so ignore all derivative terms. This is known as the local homogeneity assumption, and allows for a simple, algebraic solution. The accuracy, however, is limited by the validity of the assumption. One of the purposes of MRE is to find pathological tissue marked by a higher than normal stiffness. At the boundaries of such diseased tissue, the stiffness is expected to change, invalidating the local homogeneity assumption, and hence, the shear modulus estimate. In order to capture the true shape of any stiff regions, a method must allow for local variations.

Two new inversion methods are derived. In the first, a Green's function is introduced in an attempt to solve the differential equations. To simplify the system, the tissue is taken to be incompressible, another common assumption in direct inversion methods. Unfortunately, without designing an iterative procedure, the method still requires a homogeneity assumption, limiting potential accuracy. However, it is very fast and robust. In the second new inversion method, neither of the local homogeneity or incompressibility assumptions are made. Instead, the problem is re-posed in a quadratic optimization form. The system of linear differential equations is set as a constraint, and any free parameters are steered through quadratic programming techniques. It is found that, in most cases, there are no degrees of freedom in the optimization problem. This suggests that the system of differential equations has a fully determined solution, even without initial, boundary, or regularization conditions. The result is that estimates of the shear modulus and its derivatives can be obtained, locally, without requiring any assumptions that might invalidate the solution.

The new inversion algorithms are compared to a few prominent, existing ones, testing accuracy and robustness. The Green's function method is found to have a comparable accuracy and noise performance to existing techniques. The second inversion method, employing quadratic optimization, is shown to be significantly more accurate, but not as robust. It seems the two goals of increasing accuracy and robustness are somewhat conflicting.

One possible way to improve performance is to gather and use more data. If a second displacement field is generated using a different actuator location, further differential equations are obtained, resulting in a larger system. This enlarged system is better determined, and has improved signal-to-noise properties. It is shown that using data from a second field can increase accuracy for all methods.

Acknowledgements

I would like to thank my supervisors, Dr. Edward Vrscay and Dr. Sivabal Sivaloganathan, for accepting me as their student, for giving me the freedom to take this project wherever the wind blew, and for their support and guidance throughout my time at Waterloo. I would also like to thank Dr. Corina Drapaca for inspiring the project, for helping me get started, and for the invaluable data. Finally, thank you to all my friends, old and new, who have made these last few years enjoyable. I would not have survived Waterloo if you hadn't been there to distract me when I needed distracting, help me when I needed help, and remind me that there's more to university life than 'just school'.

To my loving parents,
Catherine Parkinson and Andrés Sánchez,
who taught me to work hard and think independently,
but also to take the time to celebrate life.

CONTENTS

List of Tables

INTRODUCTION

Stiffness in a tissue is one of the most common indicators of damage or disease. This has been known since the days of Hippocrates, when manual palpation was used regularly as a diagnostic tool [10]. It was believed that the softness or hardness of a tissue could indicate the stage of certain diseases. Today, it is common knowledge that tumours, particularly in soft tissues like the breast, are marked by stiff masses. Stiffness is also known to indicate scarring, and diseases like cirrhosis of the liver. In a paper by Manduca *et al.*, prominent researchers in the area of elastography, it is stated:

> "It is generally agreed that no other physical parameter of tissue is changed by pathological or physiological processes to as great an extent as its elasticity." [37]

This strongly suggests that the best way to find diseased tissue is to look for changes in mechanical properties.

Manual palpation, an examination by 'feeling', is extremely useful for detecting tissue maladies, but is limited to regions near the surface of the body. A physician's hand cannot feel deeply into a patient's liver, for example, without invasive surgery. Exploratory surgery can be dangerous, and is usually left as a last resort. If a method was devised to somehow 'feel' into deep tissues, non-invasively, it would be very useful diagnostic tool. This is the motivation for the development of elastography.

An object's elasticity is a measure of how it responds to physical stress. Thus, one may be able to fully characterize it by inducing stresses and measuring the corresponding response, in the form of strains, in the material. Typically, these strains are described by a set of material displacements. In elastography, measuring is performed, non-invasively, using standard imaging techniques, such as Magnetic Resonance Imaging (MRI). With the application of mathematics, the strain distribution (or displacement field) is converted to a new image called an elastogram, that maps out a parameter describing tissue stiffness. These elastograms can then be used by a clinician to aid in the detection or classification of diseased tissue. In this way, elastography can be thought of as 'palpation by imaging'.

An advantage of this technique over manual palpation is that it gives a *quantitative* description of stiffness. This allows data to be collected and compared. For example, Masuzaki *et al.* found that in seventeen patients having malignant hepatocellular carcinomas, the stiffness of the tumour was found to be near 55 kPa [42]. This can aid in diagnostics. If a new patient is found to have a mass in his liver with a stiffness value near 55 kPa, then one might infer it is a malignant tumour, so it should be tested and treated. This type of comparison is not possible with the standard qualitative palpation.

The definition of elastography is very broad. Within it, there are many classes that refer to the type of stress that is applied, the imaging modality used to measure the response, and the types of numerical techniques that transform this response into an elastogram. For example, static elastography refers to the procedure where a static (time-independent) stress is applied to the tissue. There is also quasi-static elastography, where the stress is periodically re-applied, and dynamic elastography, where the stress

is continuously changing. In ultrasound elastography, the tissue response is measured by ultrasound techniques. In magnetic resonance elastography, magnetic resonance imaging is employed. There are direct methods of inverting for the elasticity parameter, in which the data is directly used to generate an elastogram. There are also indirect methods, where the data is used as a secondary measure to test the error in an estimate. Each type of stress, imaging modality, and numerical technique has its own set of advantages and disadvantages. The 'best' method is dependent on the application.

In this thesis, only a specific class of elastography is considered. As the title suggests, it is assumed a *dynamic* stress is applied to the tissue, that the material displacement field is measured using *magnetic resonance* imaging, and that this information is *directly* used to generate a stiffness map. The focus is on the last stage, where the measured response is converted to a useable elastogram. The displacements and stiffness parameters are related through a mathematical expression, or model. Typically, these models are derived to describe the physical response of a system in terms of the parameters. In this application, the opposite is desired; a stiffness parameter, the *shear modulus*, is to be determined from a known physical response. For this reason, the task of determining the stiffness from displacements is known as an *inverse problem*. The goal is to develop new, improved inversion methods, removing some of the typical simplifying assumptions. The central thesis is that these new methods are practical, and can lead to more accurate stiffness estimates, without adding too much mathematical or computational complexity.

1.1 Outline

In the current chapter, the field of elastography is briefly introduced, and the goal of this dissertation is stated. An outline of the document is also provided, as well as a list of the most significant contributions to research contained therein.

General background related to elastography is provided in Chapter 2. This includes a description of the various classes and inversion procedures. Advantages and disadvantages of each class and procedure are listed. The existing literature concerning magnetic resonance elastography is reviewed, with particular emphasis on direct inversion techniques for harmonic (dynamic) strains.

Chapter 3 contains a ground-up development of the theory required to fully understand the problem. Many of the numerical techniques used in later chapters are developed here. The chapter is divided into four major sections. In the first, the mathematical model relating stiffness to the physical response is derived. A few of the typical simplifications of this model are also reviewed. In the second section, a full description of the MRI procedure is given. In order to capture displacements, a special encoding technique is used. While this does not directly impact the inversion methods, it does give insight into the limitations of the procedure, the type of data that is available, and what might be desired from an inversion method. The third section contains information that *is* directly applicable to the inversion. In it, general inverse problems, and standard methods to solve them, are discussed. These techniques are used in later chapters to aid in determining the stiffness parameter. The final section in Chapter 3 concerns data processing. Measured data will never be perfect, and imperfect data will always lead to imperfect results. Some of these imperfections, however, can be reduced or removed. This is the idea behind pre-processing. By 'cleaning' the data, cleaner results can be obtained. With all the theory presented in this chapter, one should fully understand the problem, and will be ready to tackle the final task of elastography: to convert displacements to stiffnesses.

An existing series of inversion methods are described, in full, in Chapter 4. These methods, known as *Algebraic Inversion of the Differential Equation* (AIDE), form the base of the novel techniques developed in future chapters. They are also the benchmark to which the new methods are compared. A small discussion is provided about some of the conclusions made in the developing works, and about how the methods might be improved.

In Chapter 5, the first novel method is developed. The idea is to use special functions, known as Green's functions, to perform the inversion. For this reason, the method is referred to as *Direct Inversion by Green's functions* (DIG). An issue arises where the Green's formulation depends on information outside the domain of measurement. To handle this, data windows are developed and introduced into the system. A series of modifications to the method are then made and compared.

A second novel inversion algorithm is developed in Chapter 6. In this method, most of the limiting assumptions made for AIDE and DIG are removed. Instead, quadratic optimization techniques are used, hence the name, *Inversion by Quadratic Optimization* (IQO). The power of this method is its potential for use with multiple data sets. It is shown that performing more scans of the same object can lead to a more accurate and stable determination of the stiffness map. Since the gathering of data can be expensive, a modification is presented, in which less data is required. Of course, this modification comes at a cost.

Chapter 7 is where most of the analysis is performed. Three data sets, to be used as test-cases, are described. Two of these are simulated, and one is from an experiment performed on a gel phantom. The methods are compared for accuracy, and performance in the presence of noise. This is, perhaps, the most visually appealing chapter, for it contains copious elastograms, side-by-side, for qualitative comparisons. The potential for improvement by pre-processing and by repeating the experiment is also explored.

In Chapters 8 and 9, final conclusions are made about the new methods, pre-processing techniques, and benefits of repeated experiments. The dissertation is summarized, highlighting key elements and results from each chapter. Possible future directions are then proposed.

While the Appendices are only meant to be supplementary, the reader may find them interesting. In the Green's function method, DIG, the continuous Green's function must be approximated by a discrete version in order to be applied to discrete data and used in digital computations. There are subtleties that arise in this discretization. These are discussed in Appendix A.

In Appendix B, a technique is developed to produce the simulated data that is used in the comparisons of Chapter 7. Generating three-dimensional displacement fields is, computationally, a very large task. Most researchers resort to simplified two-dimensional models. The new inversion methods derived in this dissertation require the full field, so this was not an option. A time-domain, explicit method is developed that uses an adapted version of Berenger's *Perfectly Matched Layer* (PML).

Finally, in Appendix C, some preliminary theory is provided about the uniqueness of solutions to this inverse problem of finding tissue stiffness from a system of differential equations. Others have incorrectly claimed that the true solution cannot be found without knowing boundary information. In the appendix, it is shown that a unique solution may exist, given only information about displacements. Furthermore, when the experiment is repeated, the enlarged differential system has a better-determined solution. This concept removes the need to make certain regularizing assumptions, so might lead to the development of further, more accurate, inversion procedures.

1.2   Contributions

Major contributions:

- Development of the *Direct Inversion by Green's Functions* (DIG) method.

    - Formal introduction of data windows into the equations of motion.
    - Derivation of smooth windows with polynomial transitions.
    - Block-inversion modification with smooth overlaps.
    - Moving-window modification.

- Development of the *Inversion by Quadratic Optimization* (IQO) method.

    - Pointwise inversion method.
    - Blockwise inversion method, with three implementations.
    - Decoupled inversion modification.

- Comparison of new methods to existing direct methods: AIDE and Optimal AIDE.

    - Analysis of accuracy and noise performance on two simulated data sets and one measured data set.
    - Analysis of possible improvement by repeating the experiment, and implications for the different methods.

Secondary contributions:

- Development of forward simulation technique to generate a three-dimensional displacement field.
    - Adaptation of Berenger's Perfectly Matched Layer to the elastic wave equation.
- Development of preliminary theory on the uniqueness of solutions.
    - Proof that it is possible for two differential equations of the type used in MRE to have a unique solution, without needing external conditions.
    - Proof that under certain conditions, four differential equations are enough to fully determine a unique solution, without needing external conditions.

There are several flavours of magnetic resonance elastography in the literature. The basic procedure for all, however, is the same: apply a known mechanical stress to the tissue; measure the physical response of the system (i.e. displacement field); and infer a quantitative description of the material's elasticity based on a mathematical model. The various methods differ in the type of applied stress, the dimensionality of the data, the assumptions made in deriving the mathematical model, and the algorithms used to estimate the elasticity parameters in these models.

In this dissertation, it is assumed that a dynamic, harmonic stress is applied, and that the three-dimensional displacement field is measured at a sequence of points in time. The tissue is modelled as a linearly viscoelastic, isotropic continuum. The parameter of interest is a complex version of the shear modulus (Lamé's second parameter). The major focus is on methods that estimate the shear modulus directly from the equation of motion, although indirect methods are also briefly reviewed.

2.1 Dynamic Elastography

There are two major types of stress that can be applied to a tissue in order to generate a physical response: static and dynamic. In static-stress elastography, a constant force is applied to the boundary of the tissue, generating fixed internal stresses. The material is imaged before and after the stress is applied, and the corresponding displacement field is computed. The Young's modulus (or some other equivalent parameter of elasticity) is estimated from a time-independent tissue model, using the static strain data. This type of experiment was performed by Skovoroda *et al.* [63], although their strain images were measured using ultrasound imaging, not MRI. In place of a strictly static stress, low frequency forcing (0–10 Hz) has also been applied in order to generate approximately static internal stresses [55]. This is known as quasi-static compression, and uses the same principles as the static case to generate an elastogram.

The major disadvantage of static elastography is that complete information about the stresses and strains at the boundaries of the region of interest must be known in order to obtain a unique solution for the stiffness [3]. Often, this boundary information must be estimated. Because of the dependence on the boundary data, the elasticity modulus estimations are *global*. That is, in order to estimate the intensity of the elastogram at a particular point, strain data from the entire volume is required. This results in the need to solve large systems of coupled equations, which requires large amounts of computing power. A second major disadvantage is that static elastography cannot estimate any time-dependent properties, such as viscous parameters. These disadvantages are somewhat overcome in dynamic elastography.

In dynamic-stress elastography, the tissue is excited in a time-dependent manner. This can either be in the form of a series of pulses [43], or harmonic forcing [41, 46]. It is the latter that is considered in this dissertation. In harmonic forcing, the surface of the tissue is mechanically excited with a periodic shear force at frequencies greater than 10 Hz. This lower frequency limit guarantees that the basic motion can

be considered vibratory [37]. The shear force generates elastic shear waves in the tissue, which induce periodic internal stresses. The force is usually applied by a motor, or mechanical driver, that is placed on one side of the tissue of interest. An example of such a device is described by Muthupillai *et al.* [46].
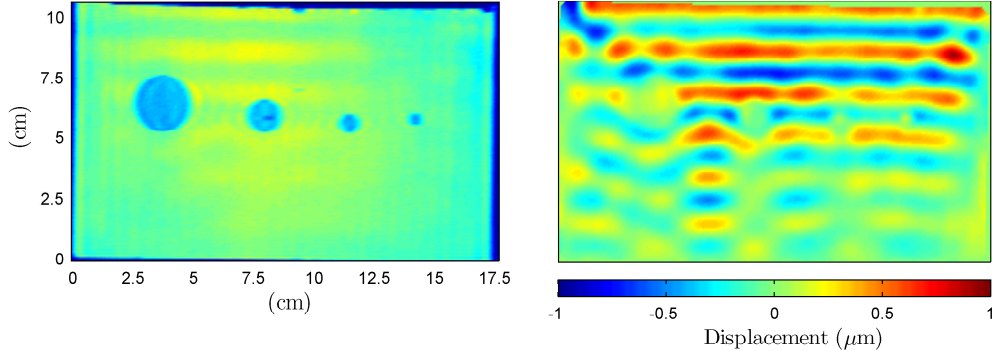
Figure 2.1: Harmonic excitation of a gel phantom. Left: MR magnitude image of a phantom containing four stiff cylindrical inclusions. Right: Displacement pattern resulting from harmonic excitation. Forcing is applied at the top of the phantom, and waves travel in the downward direction. The experiment was performed in Dr. Richard Ehman's MRE lab at the Mayo Clinic [13].

It was found that using a single mechanical driver can leave areas with low excitation due to wave attenuation and shadowing effects. Figure 2.1 shows an example of a tissue's response to harmonic forcing. In the experiment, a gel phantom was mechanically excited using a single driver. Attenuation can be seen as the waves travel downward, and reflections off the stiff inclusions cause shadow regions behind them. To compensate for these effects, multiple mechanical drivers can be used [41, 75]. By exciting the tissue from multiple positions and angles, strong patterns of displacement can be induced throughout the region of interest, leading to higher signal-to-noise ratios, and therefore improving inversion results.

With continued harmonic forcing, the internal waves eventually settle into a periodic quasi-steady state, allowing the motions to be separated into harmonics. The presence of a harmonic term in the equations of motion allows for the estimation of an elasticity parameter from *local* strain measurements [49]. This makes inversion techniques for dynamic elastography more robust (i.e. less sensitive to measurement noise) and more computationally efficient than those for static elastography. The time-dependent nature of the experiment also allows for the determination of time-dependent material properties, such as viscoelasticity. These parameters may give more insight into the nature of the examined tissue. For these reasons, the present work is focused on dynamic elastography with harmonic forcing.

Once the type of excitation is decided, an appropriate tissue model must be derived to describe the expected motions in terms of elasticity and other material parameters. The equations in the model must then be inverted, using a measured displacement field, to recover the elasticity values. This is the challenge of magnetic resonance elastography: to develop fast inversion algorithms that produce reliable elastograms. The resulting images must be suitable for use by clinicians to aid in the detection and diagnosis of pathological tissue. Existing inversion methods are reviewed in the following section.

2.2 Inversion Methods

The task of harmonic MRE is to determine an elasticity parameter from a mathematical model, given a set of measured harmonic displacements. This is a type of *inverse problem*. Inverse problems are discussed further in the next chapter. For now, it is enough to consider an inverse problem as the opposite of a well-studied forward problem. The forward problem in this instance is to determine the displacement field given the material's elasticity. There are two main classes of methods for tackling parameter estimation problems of this type: direct and indirect. In order to understand the difference between the two, consider the following system:

$$f(\mu, X) = g(X), \tag{2.1}$$

where f and g are known functionals, X is a function that can be measured, and μ is the desired parameter to be estimated. One approach to solving this system is to find the function μ that minimizes

$$E_{\text{eq}}\left(\mu, \widehat{X}\right) = \left\| f\left(\mu, \widehat{X}\right) - g\left(\widehat{X}\right) \right\|^2,$$

where $\widehat{X}$ is composed of the measured values of X. This error, E_{eq}, is known as the *equation error*. With this approach, μ is estimated *directly* from the data and Equation (2.1). For this reason, methods that minimize the equation error are known as *direct methods*.

Instead of estimating μ directly from the data, *indirect methods* begin with an initial guess of the parameter, and use it to solve for the function X that satisfies Equation (2.1). Different guesses of μ produce different outputs, $X(\mu)$. These outputs can then be compared to the measured data. The task is to search for the optimal μ that minimizes the *output error*:

$$E_{\text{out}}\left(\mu, \widehat{X}\right) = \left\| X(\mu) - \widehat{X} \right\|^2.$$

With this approach, μ is determined *indirectly*; it is decided by the solution $X(\mu)$ that best fits the data. Indirect methods often use a gradient descent or Gauss-Newton technique to search for the minimizing parameter.

Both direct and indirect methods have been applied to elastography. Direct methods often have the advantage of being faster. However, they can be sensitive to noise, and usually rely on a few limiting assumptions. Indirect methods, on the other hand, tend to be very robust. Their disadvantage is that they are always iterative. As a result, they are generally slow and rely on convergence, which is not always guaranteed without some regularizing assumptions.

2.2.1 Estimation of the Shear Modulus

All dynamic MRE methods begin with system of equations that describes motions in the tissue. An example of such a system for three-dimensional harmonic motions is the following:

$$-\rho\,\omega^2 \mathbf{U} = \nabla\cdot\left[\mathcal{M}\left(\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right)\right] + \nabla\left[\Lambda\,\nabla\cdot\mathbf{U}\right], \tag{2.2}$$

7

where ρ is the density of the material, ω is the frequency of excitation, $\mathbf{U}$ is the harmonic displacement field, $\mathcal{M}$ is the desired complex shear modulus, and Λ is the complex version of Lamé's second parameter. A detailed derivation of this system is provided in Section 3.1. Comparing Equations (2.1) and (2.2), it can be seen that elastography is a parameter estimation problem, where f is a differential operator. This adds to the complexity of the problem, since the shear modulus might not have a unique solution without specifying boundary conditions. In fact, this is an issue that is not well addressed by existing inversion methods. In Equation (2.2), the parameter $\mathcal{M}$ is defined through a system of differential equations. If this system is underdetermined, then the solution is unique only up to the addition of a function H, that satisfies the homogeneous problem:

$$\nabla \cdot \left[H \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] = 0.$$

To isolate a particular solution, external conditions, such as boundary conditions, need to be applied. Since the experiment is designed to be completely non-invasive, boundaries for the shear modulus around the region of interest cannot be directly measured. The requirement for boundary conditions also makes the parameter estimation a non-local problem; the value of the shear modulus at a particular point is dependent on information not in the vicinity of that point. This presents problems for both direct and indirect methods. Direct methods typically avoid this issue by making approximations in the equation of motion. Indirect methods use regularizing conditions, such as minimizing the total variation, that allow the iterative schemes to converge to a single solution.

2.2.1.1   Indirect Methods

In order to calculate the output-error, the forward problem must be solved given an estimate of the parameter of interest. The forward simulation of a full three-dimensional displacement field is, in itself, a daunting task; displacements oriented in each of the three orthogonal directions are coupled, and the grid dimensions can be prohibitively large. In an indirect method, this forward calculation is not only performed once, but must be repeated every time the parameter estimates are updated. This can lead to extremely long computation times. For this reason, many groups make simplifying assumptions.

One of the most common assumptions, particularly in ultrasound elastography, is that the strain waves remain in-plane [6, 11, 17, 26, 36]. This reduces the problem to two-dimensions, which is much more computationally feasible, but comes at the cost of accuracy. Since most ultrasound scanners only measure two-dimensional information along an acoustic scanning plane, the simplifying assumption is quite convenient. For these two-dimensional methods, the forward problem was either simulated using finite elements [17, 26], or an iterative Green's function technique [36]. Once the output-error was determined, parameters were updated using either a modified gradient-descent technique, or Newton's method. The plane-wave assumption, unfortunately, counteracts one of the major advantages of obtaining displacements by MRI. Magnetic resonance imaging has the ability to capture a detailed three-dimensional displacement field, which should be taken advantage of in an inversion procedure.

The first group to successfully tackle 3D elastography with an indirect iterative scheme was Van Houten *et al.* [68]. In an attempt to reduce the computational complexity, the forward problem was solved on small interior domains, dubbed *subzones*. These subzones tile the large domain, allowing for the wanted parameter to be assembled like pieces of a puzzle. Boundary conditions for the forward problem around each subzone were taken to be the measured displacements at these points, and the

8

estimate of $\mathcal{M}$ was updated using a Newton-Raphson method. The equation Van Houten *et al.* used is the following:

$$-\rho\,\omega^2\mathbf{U} = \nabla\cdot[\mathcal{M}\nabla\mathbf{U}] + \nabla\left[(\Lambda + \mathcal{M})\,\nabla\cdot\mathbf{U}\right],$$

which does not match the true wave equation from the standard continuum mechanics derivation, Equation (2.2). The two are only equivalent if it is assumed $\mathcal{M}$ has a zero gradient, which is often done in regularization. Despite this inconsistency, the results seem promising [69–71]. This iterative technique has the advantages of being parallelizable and robust. In the first version of the method, the density was assumed constant across the domain, and the longitudinal term, involving Λ, was ignored. The method was later modified to simultaneously estimate Λ, ρ, and $\mathcal{M}$ [72]. It should be stressed, however, that more free parameters in the system requires more data to accurately determine them. A single displacement field from an experiment may not be sufficient. Without additional information, further assumptions must be made about the parameters in order for the method to converge to a solution. Convergence does not imply accuracy. If the mathematical model accurately describes the tissue, then there are theoretical 'true' solutions for the parameters. The estimate from an inversion method will only be near the true solution if the assumptions are well-founded.

For clinical applications of elastography, one could easily argue that accuracy is much more important than speed of computation. Solving the full three-dimensional forward problem repeatedly, however, can be forbidding, unless the time and computing power are available. Direct methods do not suffer this computational disadvantage, so have the potential to be both fast and accurate.

2.2.1.2   Direct Methods

Direct methods are much more popular in the literature because they are faster, and often much simpler than indirect methods. In fact, the original MRE paper, by Muthupillai *et al.* [46], introduced a direct method. It was assumed that the shear modulus of the material is related to the spatial wavelength of the shear waves, λ, through

$$\mathcal{M} = \frac{\rho\,\omega^2\lambda^2}{4\pi^2}. \tag{2.3}$$

This relation is based on a simplified model, in which it is assumed the tissue is incompressible, that there are no longitudinal waves, and that the shear modulus and density are constant across the domain. With these assumptions, motions are described by a simple system of Helmholtz equations:

$$-\rho\,\omega^2\mathbf{U} = \mathcal{M}\,\nabla^2\mathbf{U}.$$

Both sides are transformed into the frequency domain with a Fourier transform,

$$\begin{aligned}
-\rho\,\omega^2\,\mathcal{F}(\mathbf{U},\boldsymbol{k}) &= \mathcal{M}\,\mathcal{F}\left(\nabla^2\mathbf{U},\boldsymbol{k}\right) \\
&= -\mathcal{M}\,\boldsymbol{k}\cdot\boldsymbol{k}\,\mathcal{F}(\mathbf{U},\boldsymbol{k}),
\end{aligned}$$

which yields the final relation:

$$\mathcal{M} = \frac{\rho\,\omega^2}{\|\boldsymbol{k}\|^2} = \frac{\rho\,\omega^2\lambda^2}{4\pi^2}.$$

Here, $\mathcal{F}$ denotes the Fourier transform applied to the spatial dimensions. In this method, the shear modulus is very simple to compute; one only needs to estimate the spatial frequency (i.e. wavenumber) or wavelength, and the shear modulus is determined by a simple division. Unfortunately, if the modulus is spatially-dependent, the Fourier transform produces a convolution equation:

$$\mathcal{F}(\mathcal{M}, \boldsymbol{k}) * \mathcal{F}(\mathbf{U}, \boldsymbol{k}) = \frac{\rho\,\omega^2}{\|\boldsymbol{k}\|^2}\,\mathcal{F}(\mathbf{U}, \boldsymbol{k}).$$

Here, any spatial variation in $\mathcal{M}$ is spread across the entire domain, making the method ill-suited for MRE. Despite this drawback, Klatt *et al.* used this Fourier transform method to estimate hepatic stiffness in volunteers [28]. The accuracy was only tested with simulations on homogeneous samples, but when applied to the clinical trials, the method seemed to show differences in hepatic stiffness.

Manduca *et al.* resolved the issue by estimating *local* spatial frequencies [33, 34]. The algorithm, named *Local Frequency Estimation* (LFE), was based on work by Knutsson *et al.* [29], who estimated local frequencies using a series of lognormal quadrature wavelets. Using this technique to perform a type of local frequency transform, and assuming the shear modulus is at least locally constant, Equation (2.3) could be used after all. Local estimates of $\mathcal{M}$ were computed everywhere, and the global, spatially-varying $\mathcal{M}$ was then constructed from all the local values, like approximating a continuous function by piecewise constants.

Instead of examining local spatial wavelengths, Bishop *et al.* estimated the shear modulus by measuring temporal wave speed in the tissue [7]. By assuming that the tissue was incompressible and had a locally constant shear modulus, the time-dependent equation of motion for the elastic medium reduced to a wave equation:

$$\frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \mathcal{M}\,\nabla^2 \boldsymbol{u},$$

where $\boldsymbol{u}$ is the time-dependent displacement field. This is a simple wave equation with wave speed $\sqrt{\mathcal{M}}$. Harmonic shear waves were sent through the tissue, and a one-dimensional profile was estimated by averaging displacements along a direction perpendicular to wave propagation. This requires the additional assumption that the shear waves travel along a single plane. Several one-dimensional profiles were created at a sequence of times, and were used to determine the speed of propagation. The wave speed was squared to give an estimate of the shear modulus.

This technique of estimating the shear modulus through wave speed was also used by McCracken *et al.* [43]. Instead of harmonic excitation, single-period wave pulses were sent through the tissue, and the transient response was measured. Images were taken with differing time-offsets from the pulse, and the time-of-arrival (TOA) was tracked, allowing for a local estimation of the wave speed. This removed the assumption of plane-wave motion, but still required incompressibility and local homogeneity.

Romano *et al.* [57] were the first to propose a method that did not depend on local wavelengths or wavespeeds. Instead, the original system of equations was transformed into a variational form, which was used to directly estimate the elasticity parameters. Multiplying the harmonic equations by a test

function, $\boldsymbol{v}$, and integrating over the volume, the following is obtained:

$$\iiint\limits_{\Omega} \left(\rho\,\omega^2\mathbf{U} + \nabla\cdot\left[\mathcal{M}\left(\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right)\right] + \nabla\left[\Lambda\nabla\cdot\mathbf{U}\right] \right) \cdot \boldsymbol{v}\,d\boldsymbol{x} = 0,$$

where Ω is the volume of interest. A sequence of test functions was chosen, $\{\boldsymbol{v}_i\}$, each having support in a small region, Ω_{loc}. The unknown parameters were assumed constant over this small domain, which results in the algebraic expression:

$$\frac{\mathcal{M}}{\rho} \underbrace{\iiint\limits_{\Omega_{\mathrm{loc}}} \boldsymbol{v}_i \cdot \left(\nabla^2\mathbf{U} + \nabla\nabla\cdot\mathbf{U}\right) d\boldsymbol{x}}_{=A_{i1}} + \frac{\Lambda}{\rho} \underbrace{\iiint\limits_{\Omega_{\mathrm{loc}}} \boldsymbol{v}_i \cdot \nabla\nabla\cdot\mathbf{U}\,d\boldsymbol{x}}_{=A_{i2}} = - \underbrace{\iiint\limits_{\Omega_{\mathrm{loc}}} \omega^2\boldsymbol{v}_i \cdot \mathbf{U}\,d\boldsymbol{x}}_{=R_i}$$

$$\frac{\mathcal{M}}{\rho}\,A_{i1} + \frac{\Lambda}{\rho}\,A_{i2} = R_i.$$

For a smooth base element, v, three test functions were chosen: $\boldsymbol{v}_1 = (v,0,0)$, $\boldsymbol{v}_2 = (0,v,0)$ and $\boldsymbol{v}_3 = (0,0,v)$. The least-squares solution to this overdetermined system is

$$\begin{bmatrix} \mathcal{M}/\rho \\ \Lambda/\rho \end{bmatrix} = \left(A^{\mathrm{H}}A\right)^{-1} A^{\mathrm{H}}R,$$

where $A = [A_{i1}\ A_{i2}]$, and A^{H} is the conjugate transpose of A. In this way, a local estimate of the elasticity-to-density ratio was computed. By repeating this procedure while shifting the support of $\boldsymbol{v}$, a global estimate was pieced together. This was the first of the algebraic inversion methods.

The next matrix method for estimating the shear modulus was introduced by Oliphant *et al.* [50]. The differential form of the equation was used, with the assumption that Λ and $\mathcal{M}$ were approximately constant, at least locally:

$$\mathcal{M}\nabla^2\mathbf{U} + (\Lambda + \mathcal{M})\,\nabla\nabla\cdot\mathbf{U} = -\rho\,\omega^2\mathbf{U}$$

$$\begin{bmatrix} \nabla^2\mathbf{U} & \nabla\nabla\cdot\mathbf{U} \end{bmatrix} \begin{bmatrix} \mathcal{M} \\ \Lambda + \mathcal{M} \end{bmatrix} = -\rho\omega^2\mathbf{U}.$$

This is an overdetermined system of equations, and can be solved using least-squares methods. The method, which was named *Algebraic Inversion of the Differential Equation* (AIDE), was later improved by Oliphant with a series of modifications [49, 52]. These are discussed in detail in Chapter 4. The drawback of the original AIDE method is that it involves high-order derivatives of data that will contain measurement noise. Differentiation amplifies high-frequency noise, making this method very sensitive. In Romano's variational formulation, integration averages the noise over the support of the test function. The technique presented here instead relies on filtering techniques to pre-process the data.

Park *et al.* [54] returned to the variational formulation. Instead of being treated as piecewise con-

stants, the unknown parameters were expanded in terms of a finite element basis:

$$U_j(\boldsymbol{x}) = \sum_{i=1}^{N_u} u_{ji}\psi_i(\boldsymbol{x}) \qquad \mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{N_\mu} \mu_i\hat{\psi}_i(\boldsymbol{x}) \qquad P(\boldsymbol{x}) = \sum_{i=1}^{N_p} p_i\breve{\psi}_i(\boldsymbol{x}),$$

where P is a pressure term that replaces the longitudinal component, $\Lambda\nabla\cdot\mathbf{U}$. Using the finite element basis functions as a set of test functions, and performing the corresponding integrations, a system is generated of the following form:

$$K(\boldsymbol{u})\,\boldsymbol{\mu} + G\,\boldsymbol{p} = \boldsymbol{f}(\boldsymbol{u}),$$

where K and G are coefficient matrices, and f is a forcing vector. From this, the shear modulus and pressure can be determined using least-squares. Unlike all previous direct methods, the local homogeneity assumption is not invoked. Instead, boundary conditions are naturally imposed by the finite element formulation. The problem of solving a system partial differential equations without knowing the boundary conditions is not avoided; the determined solutions can still be modified by any solution to the homogeneous equation.

Since the inverse problem is already ill-posed due to a lack of information, most studies in MRE restrict the tissue model to be isotropic; that is, there is no preferred direction of motion. This significantly reduces the number of free parameters in the stress-strain relationship. Anisotropic models have also been attempted. Sinkus *et al.* [59, 60] introduced a cylindrically symmetric anisotropic model, using a symmetric Young's Modulus, E. The tissue parameters were assumed locally constant, and the following equation was employed:

$$\left(-\rho\omega^2 - \iota\omega\gamma\right)U_i = \frac{E_{jj}}{2(1+\nu)}U_{i,jj} + \frac{E_{ij}}{2(1+\nu)(1-2\nu)}U_{j,ji},$$

where ν and γ are Poisson's ratio and a damping coefficient, respectively, which were assumed to be known. This Young's Modulus is composed of six free parameters at every point in space, which requires at least six equations to solve. To generate enough data, it was assumed that the equation is satisfied separately by each of the real and imaginary components of harmonic displacements. A potentially determined (but not overdetermined) system was thus created, and all six unknowns were solved algebraically. This method is incredibly sensitive to measurement noise because there is no averaging, either in the form of integration, like the variational methods, or by least-squares, like the AIDE technique. It also requires the knowledge of material parameters, which will likely not be known in practice.

The direct methods described in this section all estimate elasticity directly from the data and the equations of motion. This was either accomplished by relating the shear modulus to a local wavelength or wave speed, or by approximating the original equations by an algebraic system. In either case, the inversion process was rather simple and non-iterative, allowing for a rapid estimation of the desired parameters.

2.2.2   Summary

The goal of elastography is to determine a map of tissue elasticity based on a displacement field and a mathematical model. This is a parameter estimation problem in which elasticity is related to displace-

ments through a system of partial differential equations. The various classes of elastography differ in the type of stress that is applied, the imaging modalities used to measure the physical response of the tissue, and the form of algorithms used to convert strain images to useable elastograms.

Two types of forcing were discussed to induce strains in the material: static and dynamic. In static forcing, a constant stress is applied to the tissue. This has the disadvantage that elasticity inversion is global, and can be sensitive to boundary information. Also, time-dependent properties, such as viscoelasticity, cannot be estimated using a static model. In dynamic forcing, a time-dependent stress is applied to the tissue. The most common type is harmonic forcing, which leads to a harmonic displacement field and harmonic equations of motion. If the desired parameters are assumed to be locally constant, then the the harmonic terms in the equations of motion allow for local inversions. This results in more robust and computationally feasible inversion methods.

Given a type of forcing and tissue model, there are two classes of inversion methods: indirect and direct. Indirect inversion methods are always iterative. An initial estimate of the desired parameters is required, and the forward problem is solved. The output of this forward problem is then compared to the data, and the error is used to steer the parameters in an iterative update procedure. The best-fitting parameters are defined as those that minimize the error between the measured data and the output of the forward problem (output error). Direct methods, on the other hand, estimate the parameters directly from the data and the mathematical model. The best-fitting parameters are defined as those that minimize the error in the equations (equation error).

Outlines of the existing inversion methods were presented in this chapter. Due to the computationally difficult task of solving the full forward problem, the indirect methods either make use of a planar-wave assumption, or solve the problem on small sub-domains. This forward problem must be solved at each parameter update iteration, making indirect methods slow. For this reason, most groups attempt direct inversions. Direct methods can be separated into two classes: those that use an estimated wave speed or spatial wavelength to approximate the shear modulus, and those that use simplified versions of the equations of motion to solve for the unknown parameters algebraically. These direct methods have the advantage of being fast, but can be sensitive to noise. To compensate, pre-processing techniques and regularization should be employed.

Now that an overview of the experiment and existing inversion techniques has been provided, the specifics can be discussed in complete detail. In the next chapter, all background theory pertaining to the mathematical model, data acquisition, general inverse problems, and data pre-processing is discussed. In the chapters that follow, specific inversion methods are developed and compared.

W hen a force is applied to an object, there are physical laws that govern how the material will respond. These can often be described by simple mathematical expressions, involving few material parameters. In this dissertation, the tissue is modelled as an isotropic, linearly viscoelastic continuum, which creates an implicit relationship between material displacements and parameters that describe elasticity. In the MRE experiment, a harmonic shear force is applied to the boundary of the tissue, inducing shear waves. These are captured by a sequence of motion-encoding magnetic resonance scans. The result is a set of discrete measurements from a time-dependent three-dimensional displacement field. The task is then to estimate the parameters in the tissue model using the measured data. This is a typical example of an *inverse problem*, which may require regularization methods in order to determine a particular solution. In practice, the data will always contain noise or other forms of unwanted information. Pre-processing methods are often necessary to 'clean' the data, transforming it into a useful form that better fits the model.

In this chapter, all background theory pertaining to the tissue model, the data acquisition technique, inverse problems in general, and data pre-processing, is presented. This theory sets the stage for the inversion methods developed in the remainder of this thesis.

3.1 Equations of Motion

In order to apply the classical theory of continuum mechanics, the soft tissue is modelled as a continuum with spatially varying properties. Since the displacements are very small in the dynamic elastography experiment (on the order of a micrometre), a linear viscoelastic stress-strain relationship is assumed. The tissue is also assumed to be isotropic, meaning that there is no preferred direction of motion; the system is locally invariant to rigid spatial rotations. From these assumptions, the classical linear viscoelastic wave equation is derived for a three-dimensional continuum. It is from this equation that the elasticity parameters are estimated in most dynamic MRE inversion techniques.

3.1.1 Lagrangian versus Eulerian

The equations of motion in a continuum are classically written in one of two spatial coordinate systems: *Lagrangian* or *Eulerian coordinates*. In a Lagrangian coordinate system (sometimes called *material* coordinates), motion is described from the point of view of a 'particle' moving through space. In contrast, an Eulerian coordinate system describes motion as a function of spatial position. A Lagrangian description is perhaps more useful in the context of MRE, since properties of the tissue, such as elasticity, will typically follow the material as it moves.

Consider a region of material, Ω_o, that undergoes a translation and deformation such as the one shown in Figure 3.1. A 'particle' that begins at position $\boldsymbol{x} = \boldsymbol{a}$ is moved to the new position $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{a}, t)$
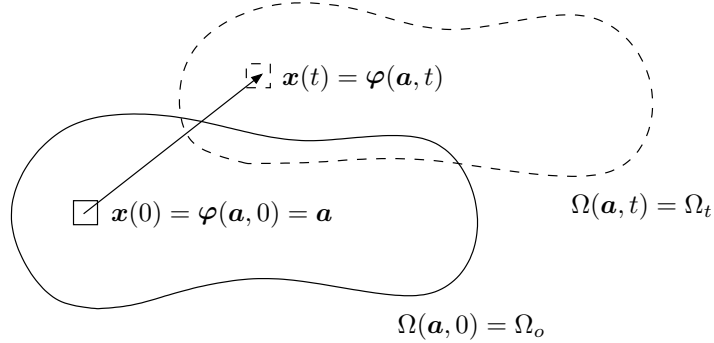
Figure 3.1: Lagrangian versus Eulerian coordinates. The object has undergone a translation and deformation. Properties of the material in its new configuration can be described in terms of spatial coordinates (Eulerian) or in terms of its original configuration (Lagrangian).

at time t. The function $\boldsymbol{\varphi}$ describes the flow of the material in time. All motions are assumed to be continuous and invertible. Any property of the material can either be described as a function of space, or as a function of the material's initial configuration. The latter may be more useful, since some properties are inherent to each 'particle' and will follow the material. The two descriptions are related:

$$f^E(\boldsymbol{x}, t) = f^E(\boldsymbol{\varphi}(\boldsymbol{a}, t), t) = f(\boldsymbol{a}, t),$$

where a superscript E denotes the Eulerian description, and f is any function describing the material. When no superscript is provided, a Lagrangian description is assumed. Note that the domains of the two functions are different. The spatial domain of f^E is the region Ω_t, while the spatial domain of f is the region Ω_o. Although there is a fundamental difference between the two descriptions, the difference in values between f^E and f can be negligible for small deformations. This will be the case when considering internal stresses in the MRE experiment.

3.1.2 Stresses and Strains

When a force is applied to the boundary of a pliable material, that force causes the material to deform. The force per unit area on the boundary is referred to as a *stress*, and the deformation can be expressed by a dimensionless quantity known as the *strain*.
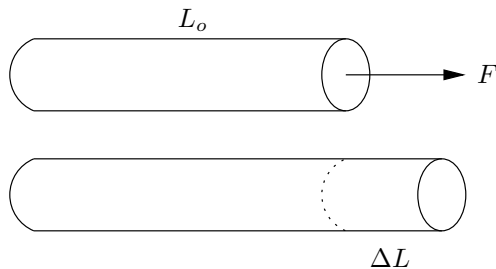
Figure 3.2: A rod acted on by a tensile force, F, is stretched to a new length $L = L_o + \Delta L$.

15

Consider a rod that is acted on by a force in the axial direction (Figure 3.2). The stress, σ, is defined as the force per unit area acting on the end of the rod. The corresponding strain is usually defined in one of two ways:

$$E(L, L_o) = \frac{L^2 - L_o^2}{2L_o^2} \qquad \text{(Euler-Almansi strain)}$$

$$\epsilon(L, L_o) = \frac{L - L_o}{L_o} \qquad \text{(Cauchy strain)}$$

Strain is a unitless parameter that describes deformation. The Cauchy strain is simply the linearized version of the Euler-Almansi strain:

$$E(L, L_o) \approx E(L_o, L_o) + \frac{\partial E}{\partial L}(L_o, L_o)[L - L_o]$$
$$= \frac{L - L_o}{L_o}$$
$$= \epsilon(L, L_o).$$

These descriptions only apply to a one-dimensional strain. For three-dimensional strain, consider a coordinate transformation,

$$(x_1, x_2, x_3) \longrightarrow (x_1', x_2', x_3').$$

The infinitesimal version of the Euler-Almansi strain tensor is given by

$$E = \frac{(dl')^2 - (dl)^2}{2(dl)^2},$$

where the infinitesimal lengths are computed as follows:

$$(dl)^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2,$$
$$(dl')^2 = (dx_1')^2 + (dx_2')^2 + (dx_3')^2.$$

Let $u_i = x_i' - x_i$ be the displacement in the ith direction. Since the new coordinates, x_i', are functions of the original coordinates, x_i, the transformed infinitesimals can be expressed in terms of the originals:

$$dx_i' = x_{i,j}' dx_j$$
$$= (u_{i,j} + x_{i,j}) dx_j$$
$$= u_{i,j} dx_j + dx_i,$$

where indices after a comma denote a partial derivative, and repeated indices indicate summation. The

total strain then becomes

$$
\begin{aligned}
E &= \frac{(dx_i dx_i + 2\, u_{i,j} dx_i dx_j + u_{i,j} u_{i,k} dx_k dx_j) - (dx_i dx_i)}{2\, dx_m dx_m} \\
&= \frac{u_{i,j} dx_i dx_j + u_{j,i} dx_i dx_j + u_{k,j} u_{k,i} dx_i dx_j}{2\, dx_m dx_m} \\
&= \frac{\frac{1}{2} \left(u_{i,j} + u_{j,i} + u_{k,i} u_{k,j} \right) dx_i dx_j}{dx_m dx_m}.
\end{aligned}
$$

From this expression, a strain tensor is defined:

$$
E_{ij} = \frac{1}{2} \left(u_{j,i} + u_{i,j} + u_{k,i} u_{k,j} \right). \tag{3.1}
$$

This is known as the Green-St. Venant tensor. The quadratic terms are often cumbersome to work with. For small displacements, the relation can be linearized, resulting in the Cauchy stress tensor:

$$
\varepsilon_{ij} = \frac{1}{2} \left(u_{j,i} + u_{i,j} \right). \tag{3.2}
$$

When combined with an appropriate constitutive equation, this expression allows the internal stresses of a material to be written as a linear function of the displacements.

3.1.3 Conservation of Momentum

The continuum equation of motion is derived from Newton's second law, which states: the rate of change of momentum in time is equal to the net force acting on the system. The momentum of an infinitesimal volume is given by

$$
\rho(\boldsymbol{a}, t) \boldsymbol{v}\left(\boldsymbol{a}, t\right) = \rho^E(\boldsymbol{x}, t)\, \boldsymbol{v}^E(\boldsymbol{x}, t),
$$

where $\boldsymbol{v}$ is the velocity. In Lagrangian coordinates, $\boldsymbol{v} = \boldsymbol{\varphi}_{,t}$. The forces acting on a volume element consist of body forces, and a surface force known as traction. Body forces are those that are proportional to mass, such as gravity. Traction comes from the interactions between a volume element and the surrounding continuum. Cauchy's principle states that the presence of material outside a volume element must exert a force on that element. This interaction can be interpreted as a force acting on the surface of the element, denoted $\boldsymbol{t}(\boldsymbol{x}, t, \hat{\boldsymbol{n}})$, where $\hat{\boldsymbol{n}}$ is an outward-pointing unit vector that is perpendicular to the surface of the element. By Cauchy's fundamental theorem for stress, this traction can be decomposed:

$$
\boldsymbol{t}(\boldsymbol{x}, t, \hat{\boldsymbol{n}}) = \boldsymbol{\sigma}(\boldsymbol{x}, t) \cdot \hat{\boldsymbol{n}} = \boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{x}, t)\, \hat{\boldsymbol{n}}.
$$

Here, $\boldsymbol{\sigma}$ is the stress tensor, written in matrix form. The stress tensor is related to the previously derived strain tensor through the use of a constitutive equation. Thus, for some volume Ω_t, balancing linear momentum (Newton's second law) results in the following integral equation of motion:

$$
\frac{d}{dt} \int_{\Omega_t} \rho^E(\boldsymbol{x}, t) \boldsymbol{v}^E(\boldsymbol{x}, t) dV_t = \int_{\partial \Omega_t} \boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{x}, t)\, \hat{\boldsymbol{n}}_t dA_t + \int_{\Omega_t} \rho^E(\boldsymbol{x}, t) \boldsymbol{b}^E(\boldsymbol{x}, t) dV_t,
$$

where $\boldsymbol{b}^E$ are the body forces, $\partial\Omega_t$ is the surface of the volume, and a subscript t denotes that the coordinate system is dependent on time t. This equation can be transformed into Lagrangian coordinates,

$$\frac{d}{dt}\int_{\Omega_o}\rho(\boldsymbol{a},t)\boldsymbol{v}(\boldsymbol{a},t)\,|\boldsymbol{F}|\,dV_o = \int_{\partial\Omega_o}\boldsymbol{F}^{-1}\boldsymbol{\sigma}^{\mathrm{T}}(\boldsymbol{a},t)\,\hat{\boldsymbol{n}}_o\,|\boldsymbol{F}|\,dA_o + \int_{\Omega_o}\rho(\boldsymbol{a},t)\boldsymbol{b}(\boldsymbol{a},t)\,|\boldsymbol{F}|\,dV_o,$$

where $\boldsymbol{F}$ is the Jacobian matrix for the coordinate change, and $|\boldsymbol{F}|$ is the Jacobian determinant. The surface integral was transformed with the use of Nanson's formula. For the flow functions, $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{a},t)$, the Jacobian is given by $\boldsymbol{F} = \varphi_{i,j}$. Define the matrix $\boldsymbol{P} = |\boldsymbol{F}|\boldsymbol{\sigma}\,\boldsymbol{F}^{-\mathrm{T}}$. This is known as the *first Piola-Kirchoff stress tensor*. It relates stress to the undeformed area. With this substitution, and invoking the Divergence theorem, the integral equation becomes:

$$\int_{\Omega_o}\left[\boldsymbol{v}\,\frac{d}{dt}\left(\rho\,|\boldsymbol{F}|\right) + \rho\,|\boldsymbol{F}|\,\frac{d\boldsymbol{v}}{dt}\right]dV_o = \int_{\partial\Omega_o}\boldsymbol{P}^{\mathrm{T}}\hat{\boldsymbol{n}}_o dA + \int_{\Omega_o}\rho\boldsymbol{b}\,|\boldsymbol{F}|\,dV_o$$

$$\int_{\Omega_o}\left[\boldsymbol{v}\,\frac{d}{dt}\left(\rho\,|\boldsymbol{F}|\right) + \rho\,|\boldsymbol{F}|\,\frac{d\boldsymbol{v}}{dt}\right]dV_o = \int_{\Omega_o}\nabla\cdot\boldsymbol{P}\,dV_o + \int_{\Omega_o}\rho\boldsymbol{b}\,|\boldsymbol{F}|\,dV_o$$

$$\int_{\Omega_o}\left[\boldsymbol{v}\,\frac{d}{dt}\left(\rho\,|\boldsymbol{F}|\right) + \rho\,|\boldsymbol{F}|\,\frac{d\boldsymbol{v}}{dt} - \nabla\cdot\boldsymbol{P} - \rho\boldsymbol{b}\,|\boldsymbol{F}|\right]dV_o = 0.$$

If it is assumed that the integrand is continuous, the DuBois-Reymond lemma can be applied. The integrand must be identically zero since the above integral must hold for all initial configurations Ω_o. Therefore, the balance of linear momentum results in the following differential equation:

$$\boldsymbol{v}\,\frac{d}{dt}\left(\rho\,|\boldsymbol{F}|\right) + \rho\,|\boldsymbol{F}|\,\frac{d\boldsymbol{v}}{dt} = \nabla\cdot\boldsymbol{P} + \rho\,|\boldsymbol{F}|\,\boldsymbol{b}. \tag{3.3}$$

This system is underdetermined, since it contains more free parameters than equations. Another equation can be added by considering conservation of mass. The total mass, following the material as it deforms, is given by

$$M = \int_{\Omega_t}\rho^E(\boldsymbol{x},t)V_t$$

Since the MRE experiment will not likely be powerful enough to convert matter into energy, and it is assumed there are no other mass sources or sinks, this total mass should remain unchanged as time progresses:

$$\frac{dM}{dt} = \frac{d}{dt}\int_{\Omega_t}\rho^E(\boldsymbol{x},t)dV_t = 0.$$

Changing to a Lagrangian coordinate system allows the time derivative to be brought inside the integral, resulting in

$$\frac{d}{dt}\int_{\Omega_o}\rho(\boldsymbol{a},t)\,|\boldsymbol{F}|\,dV_o = \int_{\Omega_o}\frac{d}{dt}\left(\rho\,|\boldsymbol{F}|\right)dV_o = 0.$$

Again, if the integrand is continuous, the DuBois-Reymond lemma can be applied:

$$\frac{d}{dt}\left(\rho\left|\boldsymbol{F}\right|\right) = 0$$

$$\implies \qquad \rho\left|\boldsymbol{F}\right| = \rho_o.$$

This is an expression for the conservation of mass in Lagrangian coordinates. It is most often seen in the equivalent Eulerian description,

$$\frac{\partial \rho^E}{\partial t} + \nabla\cdot\left(\rho^E \boldsymbol{v}^E\right) = 0.$$

This is sometimes referred to as the *continuity equation*. Combining this result with the conservation of momentum, the following system is derived:

$$\rho_o\,\frac{d\boldsymbol{v}}{dt} = \nabla\cdot\boldsymbol{P} + \rho_o\,\boldsymbol{b}. \tag{3.4}$$

Assuming small deformations, $\boldsymbol{F} \approx \boldsymbol{I}$ and $\left|\boldsymbol{F}\right| \approx 1$. Thus, the Piola-Kirchoff stress tensor is approximately equal to the Cauchy stress tensor for small motions. Making this approximation, and letting $\boldsymbol{u}(\boldsymbol{a},t) = \boldsymbol{\varphi}(\boldsymbol{a},t) - \boldsymbol{a}$, the final system for small deformations is the following:

$$\rho_o\,\frac{\partial^2}{\partial t^2}\boldsymbol{u} = \nabla\cdot\boldsymbol{\sigma} + \rho_o\,\boldsymbol{b} \tag{3.5}$$

This is actually a system of three equations, where time-dependent motions are related to the internal stresses of the material. It has already been discussed that the internal stresses are related to strains, and that these strains are related to displacements by Equation (3.2). All that is required is a constitutive equation to provide an explicit relation between the stress tensor, $\boldsymbol{\sigma}$, and the strain tensor, $\boldsymbol{\varepsilon}$. This is accomplished with the use of a linear viscoelastic model.

3.1.4   Linear Viscoelasticity

Viscoelastic materials are so-named because they exhibit both viscous and elastic characteristics. Viscous materials tend to resist flow, or strain, when a stress is applied. For example, if honey is poured onto a plate, it will initially form a blob where it lands. Over time, gravity acting on it will force the honey to slowly spread out until it forms an even layer. Water, on the other hand, is considered non-viscous, and will spread out almost instantaneously. Purely elastic materials deform instantaneously when a stress is applied. More importantly, when that stress is removed, the material instantly returns to its initial configuration. Viscoelasticity is a combination of the two: there is a time-dependent relationship between stress and strain in which the material tends to resist a change in shape, but no permanent strain is observed. Viscoelastic materials always have the ability to return to any previous state.

In a linear viscoelastic model, the relationship between stress and strain is linear. That is, the following must be satisfied:

$$\varepsilon\left[c\,\sigma(t)\right] = c\,\varepsilon\left[\sigma(t)\right],$$

$$\varepsilon\left[\sigma_1(t) + \sigma_2(t)\right] = \varepsilon\left[\sigma_1(t)\right] + \varepsilon[\sigma_2(t)],$$

where ε and σ are the strain and stress, respectively, and c is an arbitrary constant.

Typically, linear viscoelastic relationships are derived from models using various combinations of springs (which exhibit elastic behaviour) and dashpots (which exhibit viscous behaviour). The most common are the Maxwell model, which consists of a spring and dashpot connected in series, and the Kelvin-Voigt model, which consists of a spring and dashpot connected in parallel. Other specific constitutive models have been explored by McGarry *et al.* [44]. Here, the general case is considered. Regardless of the model, the relationship between stress and strain can be written in the following form:

$$\varepsilon(t) = H(t)\,\sigma(t),$$

where $H(t)$ is a linear differential or integral operator (or combination of the two). In order to write the stress as a function of strain, consider the impulse response of the system:

$$H(t)\,g(t,\tau) = \delta(t - \tau).$$

The function g is known as the Green's function for the operator H. With this Green's function, σ is determined by an integral,

$$\sigma(t) = \int_{-\infty}^{\infty} g(t,\tau)\varepsilon(\tau)d\tau.$$

Applying H to both sides, it is easily shown that the original stress-strain relationship can be recovered:

$$
\begin{aligned}
H(t)\,\sigma(t) &= \int_{-\infty}^{\infty} H(t)\,g(t,\tau)\,\varepsilon(\tau)\,d\tau \\
&= \int_{-\infty}^{\infty} \delta(t - \tau)\,\varepsilon(\tau)\,d\tau \\
&= \varepsilon(t).
\end{aligned}
$$

The integral relationship between stress and strain must satisfy two constraints in most physical systems: causality, which states that an effect cannot precede its cause; and time invariance, which states that the mechanics do not depend explicitly on the physical time, but only on the relative time between adding a stress and observing the response. Due to causality, $\sigma(t)$ cannot depend on any values of $\varepsilon(\tau)$, where $\tau > t$. Therefore, it must be that $g(t,\tau) = 0$, for $\tau > t$. Also, because of time invariance, $g(t,\tau) = g(t - \tau, 0) = g(t - \tau)$. With these conditions,

$$\sigma(t) = \int_{-\infty}^{t} g(t - \tau)\varepsilon(\tau)d\tau. \tag{3.6}$$

This is the general integral form of the relationship describing stress as a function of strain. In the

literature (e.g. [64, 66]), the integral is usually written in terms of the rate of change of strain,

$$\sigma(t) = \int\limits_{-\infty}^{t} \hat{g}(t-\tau)\frac{\partial}{\partial\tau}\varepsilon(\tau)d\tau. \tag{3.7}$$

This equation can also be derived using the principle of super-position. The two expressions are equivalent, where $\hat{g}$ is a modified version of the Green's function, g, that accounts for the added derivative. This expression is not incredibly useful, since the Green's function is unknown unless a specific spring-dashpot model is provided. A more useful general representation, however, can be derived for dynamic analysis, when the viscoelastic material experiences oscillatory loading.

3.1.5   Oscillatory Loading

For experiments where a step increase in stress is applied to a viscoelastic material, the corresponding strain response can sometimes reach a steady state faster than the time it takes to perform any measurement. If this is the case, step-stress experiments are not capable of uncovering the viscoelastic behaviour; only the time-limiting elastic behaviour can be studied. For this reason, dynamic loading has been explored extensively. In dynamic loading, the material is never allowed to reach a steady-state. This allows viscous responses to be estimated on much smaller time scales than the measurement times. One such experiment assumes that the stress (or strain) takes the form of a sinusoid. This is the case of interest for the harmonic MRE experiment.

If a sinusoidal stress is applied to the boundary of a material, periodic internal strains will be induced. Consider an internal point undergoing harmonic motion with a strain amplitude ε_o, and frequency ω:

$$\varepsilon(t) = \varepsilon_o e^{i\omega t}.$$

Equation (3.6) is used to derive the corresponding stress at this point:

$$\sigma(t) = \int\limits_{-\infty}^{t} g(t-\tau)\varepsilon_o e^{i\omega\tau}d\tau$$

$$= \varepsilon_o e^{i\omega t}\underbrace{\int\limits_{0}^{\infty} g(\tau)e^{-i\omega\tau}d\tau}_{=G(\omega)}$$

$$= G(\omega)\varepsilon(t).$$

The quantity G is known as the *complex dynamic modulus* of the material. In general, it is a complex function of frequency. If it is separated into its real and imaginary components,

$$G(\omega) = G_S(\omega) + i\,G_L(\omega),$$

then the following frequency-dependent relationship is obtained:

$$\sigma(t) = [G_S(\omega) + iG_L(\omega)]\, \varepsilon(t)$$
$$= G_S(\omega)\, \varepsilon(t) + \frac{G_L(\omega)}{\omega}\, \frac{\partial \varepsilon(t)}{\partial t}. \tag{3.8}$$

The functions G_S and G_L are known as the *storage modulus* and *loss modulus*, respectively. From Equation (3.8), it is evident that if the stress is sinusoidal, then the strain is simply a phase-shifted sinusoid with the same frequency. The phase-shift, δ, satisfies the following relation, known as the *loss tangent*:

$$\tan \delta(\omega) = \frac{G_L(\omega)}{G_S(\omega)}.$$

The same derivations can be performed assuming the strain is composed of a sum of frequencies. The resultant stress consists of a sum of sinusoids with the same frequencies, but phase-shifted according to the frequency-dependent loss tangents. These moduli are still general expressions; no specific viscoelastic model is assumed. In order to relate the moduli at different frequencies, however, a specific model must be provided.

For the previous calculations, it was assumed that stress was applied along one spatial dimension. This is easily generalized for stresses and strains in three dimensions:

$$\sigma_{ij}(t) = \int_{-\infty}^{t} g_{ijkl}(t - \tau)\, \varepsilon_{kl}(\tau)\, d\tau$$
$$= G_{ijkl}(\omega)\, \varepsilon_{kl}(t), \tag{3.9}$$

where repeated indices imply summation along that dimension (Einstein notation). According to Equation (3.2), if the displacements, u_i, are assumed to be the sum of periodic functions, then the strain tensor, ε_{ij}, is also a sum of periodic functions with the same frequencies. In Section 3.2.1, it will be discussed that the data acquisition technique requires the displacement field to be periodic; therefore, the conditions for oscillatory loading are met.

This tensor version of the complex dynamic modulus, G_{ijkl}, has 81 components. If the goal of the experiment is to completely determine this linear stress-strain relationship, then there are 81 unknowns to find for each point in space, which requires an unreasonable amount of data. This number can be drastically reduced by exploiting various symmetries.

The first symmetry is in the strain tensor, which can be easily noted in Equations (3.1) and (3.2). Since $\varepsilon_{kl} = \varepsilon_{lk}$, it must be that $G_{ijkl} = G_{ijlk}$. This reduces the number of distinct elements in G to 54.

Conservation of angular momentum can be invoked to further reduce the modulus. Consider an infinitesimal volume element. As long as this volume element is not undergoing angular acceleration, momentum is conserved, implying there are no net torques. This condition forces the stress tensor to be symmetric: it must be that $\sigma_{ij} = \sigma_{ji}$. This, in turn, implies that the complex dynamic modulus must satisfy $G_{ijkl} = G_{jikl}$. With this restriction, G may only consist of 36 distinct parameters.

A third symmetry is implied by the existence of a *strain energy potential* function. In general, the

stress of a material can be derived from a scalar energy potential,

$$\sigma_{ij} = \frac{\partial w(\varepsilon)}{\partial \varepsilon_{ij}}.$$

With the linearity assumption, the precise form of the energy potential is easily derived:

$$w(\varepsilon) = G_{ijkl}\, \varepsilon_{ij}\varepsilon_{kl}$$

Interchanging the indices ij and kl does not change the above scalar potential, so it must be that $G_{ijkl} = G_{klij}$. With this restriction, G is limited to only 21 distinct entries.

The previously mentioned symmetries exist for general linearly viscoelastic materials undergoing oscillatory loading. For most applications, it is still not practical to have twenty-one unknown parameters. Other assumptions about the material can be invoked to further reduce this number of degrees of freedom. In this dissertation, it is assumed that the soft tissue possesses the property of isotropism.

3.1.6   Isotropy

A material property is known as *isotropic* if that property is not dependent on direction. In particular, the complex dynamic modulus is isotropic if it is invariant under rigid spatial rotations. This implies that motions within the material are not preferred along any one direction. With this restriction, it can be shown that G consists of only two distinct parameters, and can be written as follows:

$$G_{ijkl} = \mathcal{M}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) + \Lambda\,\delta_{ij}\delta_{kl},$$

where δ_{ij} is the identity tensor (Kronecker delta), and Λ and $\mathcal{M}$ are the complex, frequency-dependent versions of Lamé's first and second parameters, respectively. When combined with Equations (3.2) and (3.9), the stress-displacement relationship is

$$\sigma_{ij}(t) = \mathcal{M}(\omega)\left[u_{j,i}(t) + u_{i,j}(t)\right] + \Lambda(\omega)\,\delta_{ij}\,u_{kk}(t). \tag{3.10}$$

These complex versions of the Lamé parameters are related to the real versions:

$$\mathcal{R}e\{\Lambda(0)\} = \Lambda(0) = \lambda, \qquad \mathcal{R}e\{\mathcal{M}(0)\} = \mathcal{M}(0) = \mu,$$

and the real Lamé parameters are related to the Young's Modulus, E, and Poisson's Ratio, ν, through

$$E = \mu\frac{3\lambda + 2\mu}{\lambda + \mu}, \qquad \nu = \frac{\lambda}{2(\lambda + \mu)}.$$

Therefore, if the complex Lamé parameters are fully determined, the static Young's modulus and Poisson's ratio can also be calculated.

Armed with the assumptions of linear viscoelasticity, isotropy, and oscillatory loading, a general expression can be derived that relates periodic motions to the complex Lamé parameters through a three-dimensional, periodic version of the viscoelastic wave equation.

3.1.7 Periodicity

Due to the data acquisition process, the displacement field must be periodic with some fundamental frequency, ω_o, but may contain integer harmonics of this frequency. This will be discussed further in Section 3.2.1. In such a case, displacements in the kth direction can always be written as

$$u_k(t) = \sum_{n=-N}^{N} U_{kn} e^{in\omega_o t},$$

where N is the highest significant harmonic present, and $\{U_{kn}\}$ are the complex amplitudes of the signal. In order for the displacements to be a real, it must be that $U_{kn} = U_{-kn}^*$. These coefficients are precisely the Fourier series coefficients for the signal u_k. Since the data consist of displacements at discrete times, the set $\{U_k\}_n$ can be determined by performing a Discrete Fourier Transform (DFT) on u_k, as long as the conditions for Nyquist's sampling theorem apply.

Combining the periodic expression for u_k with Equations (3.5) and (3.10), the following system is obtained:

$$\sum_{n=-N}^{N} e^{in\omega_o t} \left(\rho_o\, n^2 \omega_o^2 \mathbf{U}_n + \nabla \cdot \left[\mathcal{M}(n\omega_o) \left(\nabla \mathbf{U}_n + \nabla^{\text{T}} \mathbf{U}_n \right) \right] + \nabla \left[\Lambda(n\omega_o) \nabla \cdot \mathbf{U}_n \right] \right) = -\rho_o\, \boldsymbol{b},$$

where $\nabla^{\text{T}} \mathbf{U} = [\nabla \mathbf{U}]^{\text{T}}$. The body forces, $\boldsymbol{b}$, are likely to be constant in time. However, one can consider all periodic body forces,

$$b_k(t) = \sum_{n=-N}^{N} B_{kn} e^{in\omega_o t}.$$

Grouping the coefficients of the complex exponentials,

$$\sum_{n=-N}^{N} e^{in\omega_o t} \left(\rho_o\, n^2 \omega_o^2 \mathbf{U}_n + \rho_o\, \boldsymbol{B}_n + \nabla \cdot \left[\mathcal{M}(n\omega_o) \left(\nabla \mathbf{U}_n + \nabla^{\text{T}} \mathbf{U}_n \right) \right] + \nabla \left[\Lambda(n\omega_o) \nabla \cdot \mathbf{U}_n \right] \right) = 0.$$

This is just the Fourier series of the time-dependent continuum equation of motion. The above can only hold if the coefficients for each exponential are identically zero. Thus, the system is further reduced to

$$-\rho_o \left[n^2 \omega_o^2 \mathbf{U}_n + \boldsymbol{B}_n \right] = \nabla \cdot \left[\mathcal{M}(n\omega_o) \left(\nabla \mathbf{U}_n + \nabla^{\text{T}} \mathbf{U}_n \right) \right] + \nabla \left[\Lambda(n\omega_o) \nabla \cdot \mathbf{U}_n \right]. \tag{3.11}$$

To simplify notation, dependence on frequency will henceforth be implicit:

$$-\rho_o \left[\omega^2 \mathbf{U} + \boldsymbol{B} \right] = \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\text{T}} \mathbf{U} \right) \right] + \nabla \left[\Lambda \nabla \cdot \mathbf{U} \right]. \tag{3.12}$$

Most current elastography methods begin with Equation (3.12) and attempt to solve for the shear modulus, $\mathcal{M}$. To accomplish this, the data is often transformed, or approximations are made to simplify the inversion. Some of these techniques are now described.

3.1.8 Simplifications

In the MR scanner, the only body forces acting on the subject are gravity and small electromagnetic forces. Neither of these are expected to be harmonic; therefore, the first possible simplification is the removal of the body-force term. The next most common assumption is that the tissue density is constant throughout space. The density of most soft tissues is close to that of water, approximately 1 g/cm^3. The clinical purpose of MRE is to distinguish pathological tissue that cannot normally be distinguished using conventional techniques. If the diseased tissue has a noticeably different density from the surrounding healthy tissue, then this can be picked up by standard MRI. Without making the assumption $\rho \approx 1$, only the ratio of shear modulus to density can be computed. After these two assumptions are made, the system in Equation (3.12) becomes

$$-\omega^2 \mathbf{U} = \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^\mathrm{T} \mathbf{U} \right) \right] + \nabla \left[\Lambda \, \nabla \cdot \mathbf{U} \right]. \tag{3.13}$$

The two remaining inconveniences are the lack of boundary conditions for the shear modulus, and the presence of the unknown longitudinal parameter, Λ.

In almost all of the existing direct inversion methods, the unknown parameters are approximated as constants over local domains. This is known as the *local homogeneity* assumption. If valid, this allows $\mathcal{M}$ and Λ to be pulled out of the derivatives:

$$-\omega^2 \mathbf{U} = \mathcal{M} \, \nabla^2 \mathbf{U} + (\Lambda + \mathcal{M}) \, \nabla \nabla \cdot \mathbf{U}, \tag{3.14}$$

where the identity $\nabla \nabla \cdot \mathbf{U} = \nabla \cdot \nabla^\mathrm{T} \mathbf{U}$ is used. This identity is easily shown using tensors:

$$\nabla \cdot \nabla^\mathrm{T} \mathbf{U} = \left([U_{i,j}]^\mathrm{T} \right)_{,j} = U_{j,ij} = U_{j,ji} = \left(U_{j,j} \right)_{,i} = \nabla \nabla \cdot \mathbf{U}$$

This local homogeneity assumption is powerful. Enforcing zero gradients reduces the problem to an algebraic one, removing the need for boundary information to obtain a solution. There is no guarantee, however, that the obtained solution matches the true stiffness of the material; the approximation simply attempts to find the least-varying one out of the solution family.

Due to the large amount of water present in soft tissues, they are often considered nearly incompressible. Because of this incompressible nature, the longitudinal waves travel extremely quickly through the tissue, resulting in very small differences in displacements between neighbouring volume elements. In fact, these waves can move on a time scale faster than the measurement times of the MRI scanner, making them impossible to measure accurately. Shear waves, on the other hand, travel much slower in soft tissues. It has been found, experimentally, that there is much more variation in displacements caused by the shear component. For this reason, mechanical excitation is in the form of a shear stress, and the parameter of interest is the shear modulus.

To avoid having to determine Λ as well as $\mathcal{M}$ from the equations of motion, the longitudinal component must somehow be removed. This has been tackled in several ways:

1. Assume incompressibility [15, 49, 50].

2. Assume slow variation of longitudinal waves [37, 53].

3. Assume a fixed Poisson's ratio [59, 60].

4. Replace the longitudinal component with an unknown pressure term [54].

5. Remove the longitudinal component by applying the curl operator to the equation [12, 61].

In the first three methods, an approximation is made in order to remove the Λ term. In the last two, the equation is converted into a new form that might be easier to work with.

Soft tissues are mainly composed of water. This is the main reason why they exhibit many properties of water, such as their density and incompressibility. If the tissue is completely incompressible, then the displacement field satisfies

$$\nabla \cdot \mathbf{U} = 0.$$

This completely eliminates the longitudinal term, as desired. By expanding the shear term, another possible simplification can be seen:

$$\begin{aligned} \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] &= \mathcal{M} \, \nabla \cdot \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) + \nabla \mathcal{M} \cdot \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \\ &= \mathcal{M} \, \nabla^2 \mathbf{U} + \mathcal{M} \nabla \nabla \cdot \mathbf{U} + \nabla \mathcal{M} \cdot \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \\ &= \mathcal{M} \, \nabla^2 \mathbf{U} + \nabla \mathcal{M} \cdot \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right). \end{aligned}$$

If it is further assumed that the material is locally homogeneous, the $\nabla \mathcal{M}$ term is removed, resulting in a Helmholtz equation,

$$-\omega^2 \, \mathbf{U} = \mathcal{M} \, \nabla^2 \mathbf{U}.$$

In this form, the equations of motion are decoupled. This potentially allows for an estimate of the shear modulus from a single polarization of motion. If only motions oriented in a single direction are captured, the total scan time is reduced by a factor of three. Unfortunately, in reality, soft tissues are not completely incompressible. Since the longitudinal waves travel at very fast speeds, Λ tends to be rather large, balancing out the small divergence.

The next commonly made assumption is that the entire term involving Λ can simply be ignored. The justification is that since the waves travel very quickly, the local variation in displacement caused by the longitudinal component is negligible. Removing the term results in the equation

$$-\omega^2 \mathbf{U} = \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right]. \tag{3.15}$$

Since the mechanical excitation is in the form of shear motion, this might be justified. Although the shear excitation will also generate longitudinal waves, these waves are likely to have much smaller amplitudes. Since the scanner is slower than these longitudinal waves anyway, it is hoped that the contribution of Λ will appear as a small noise component.

Instead of attempting to provide reasons why the term involving Λ can be ignored, some investigators assume it is related to the shear modulus through Poisson's ratio,

$$\nu = \frac{\lambda}{2(\lambda + \mu)} \qquad \Longrightarrow \qquad \lambda = \frac{2\nu}{1 - 2\nu} \mu,$$

where ν is Poisson's Ratio for the tissue, and λ and μ are the purely real versions of Lamé's parameters.

A constant ν is then assumed throughout the tissue with a value between 0.48 and 0.4999. If the same relationship is assumed for the complex, viscoelastic versions of Lamé's parameters, then $\Lambda \propto \mathcal{M}$:

$$-\omega^2 \mathbf{U} = \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] + \frac{2\nu}{(1-2\nu)} \nabla \left[\mathcal{M} \nabla \cdot \mathbf{U} \right].$$

The problem with this assumption is that the equation is very sensitive to ν. If it happens that the calculated divergence of displacement is not near zero, perhaps due to measurement noise, a Poisson's Ratio near 0.5 will amplify this term. This would emphasize the longitudinal component of the wave over the shear component, which is invalid.

The standard method in fluid mechanics for treating this balance between the large longitudinal parameter and small divergence of displacements is to replace $\Lambda \nabla \cdot \mathbf{U}$ with a pressure term:

$$-\omega^2 \mathbf{U} = \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] + \nabla P,$$

where P is an unknown pressure. This method is rarely used in elastography because it just replaces one unknown parameter with another. However, the form is much simpler. Park *et al.* made this approximation in their finite element model, and attempt to recover the pressure instead of Λ [54].

The final method for handling the longitudinal term involves no assumptions. Instead, the equation of motion is transformed by applying the curl operator to both sides:

$$\begin{aligned} -\omega^2 \nabla \times \mathbf{U} &= \nabla \times \nabla \left[\Lambda \nabla \cdot \mathbf{U} \right] + \nabla \times \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] \\ -\omega^2 \nabla \times \mathbf{U} &= \nabla \times \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right], \end{aligned} \qquad (3.16)$$

since $\nabla \times \nabla f = 0$ for smooth functions. The disadvantages of this method are subtle. The original problem is a first-order partial differential equation for $\mathcal{M}$, whereas the above is second-order. Without boundary conditions, the family of solutions is larger for this new problem. Solutions to Equation (3.13) are unique up to the addition of a function, H_1, that satisfies

$$\nabla \cdot \left[H_1 \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] = 0,$$

whereas solutions to Equation (3.16) are unique up to the addition of a function, H_2, that satisfies

$$\nabla \times \nabla \cdot \left[H_2 \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] = 0.$$

The solution space to this second equation is larger than and contains the first. This means that if one obtained a solution to Equation (3.16), it may not satisfy the original equations of motion. It was already put forth that a solution to Equation (3.13) may not match the true distribution of the shear modulus, so applying the curl weakens the result even further. The other disadvantage is that the local homogeneity assumption now implies that all the second derivatives of $\mathcal{M}$ are also zero. This limits the allowable variation even further. Wherever the assumption breaks down, the stiffness estimate is invalid.

The purpose of the novel work in this dissertation is to construct inversion methods that rely on fewer assumptions. In particular, it is desired to remove the local homogeneity assumption wherever possible. The purpose of magnetic resonance elastography is to find stiff inclusions based on differences in stiffness. Assuming that the shear modulus is locally constant is counter-productive; stiff inclusions

are often best marked by regions where the modulus changes. Homogeneity smooths out these regions, and may cause the method to completely miss small inclusions.

3.1.9 Summary

In this section, the soft tissue model was derived. The model can be expressed in either Eulerian or Lagrangian coordinates. For the relatively small displacements expected in the MRE experiment, however, the two are approximately equivalent. A linear viscoelastic isotropic stress-strain relationship is adopted, which involves two parameters: Λ and $\mathcal{M}$. When this is combined with the conservation of momentum and conservation of mass equations, the three-dimensional viscoelastic harmonic wave equation is obtained:

$$-\rho_o\,\omega^2\mathbf{U} = \nabla\cdot\left[\mathcal{M}\left(\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right)\right] + \nabla\left[\Lambda\,\nabla\cdot\mathbf{U}\right].$$

It is from this equation that the stiffness parameter, $\mathcal{M}$, must be estimated.

There are two major issues in inverting the equation of motion for the complex shear modulus: $\mathcal{M}$ is expressed in a differential form with no known boundary conditions, and Λ is unknown and undesired, but must either be estimated or eliminated. The lack of boundary information can be addressed by assuming local homogeneity. This is a regularizing assumption. With it, the equation becomes algebraic in terms of the unknown parameters, which can be solved using least-squares methods. The longitudinal parameter, Λ, can be removed by either ignoring terms that are expected to be small, or by transforming the equations.

Each assumption or transformation weakens the resulting elasticity estimate. Important features might be smoothed out or removed. For example, a solid tumour might be best indicated by a sharp change in tissue elasticity. All sharp changes are smoothed out, however, if the local homogeneity assumption is made. For this reason, the goal of this dissertation is to devise methods that do not depend on these assumptions.

The harmonic tissue model was derived with the assumption that the displacement field is periodic. This property is actually required by the data acquisition method, so was built into the model. The imaging technique, as well as the governing physics behind it, is now discussed in the next section. It is important to know how the data is obtained in order to fully understand possible sources of error, as well as the reasons for certain assumptions. This knowledge is also useful when postulating ways to improve inversions that depend on the availability of data.

3.2 Magnetic Resonance Imaging

In magnetic resonance elastography, the displacement field within a tissue is measured using a magnetic resonance scanner. This has a large advantage over ultrasound elastography. The MR scanner can non-invasively penetrate deeper into tissues than an ultrasound scanner, has a higher resolution, and can easily image a full three-dimensional volume. This allows for the acquisition of more data with a higher level of detail than ultrasound techniques. As a result, the elasticity inversion methods have the potential to be more accurate. A review of the physics and methods involved in MRI is presented in this section, as well as a description of the techniques used to encode and recover motion information.

3.2.1 Data Acquisition

The amount of data required by the MRE experiment is rather large. The equations of motion derived in this chapter involve a time-dependent displacement field over a three-dimensional volume. For the experiments considered in this dissertation, the spatial grid has dimensions $256 \times 256 \times 16$. This results in a need for $3\,145\,728$ displacement values to be measured at a given instant in time. Not surprisingly, limitations of the scanner prevent this.
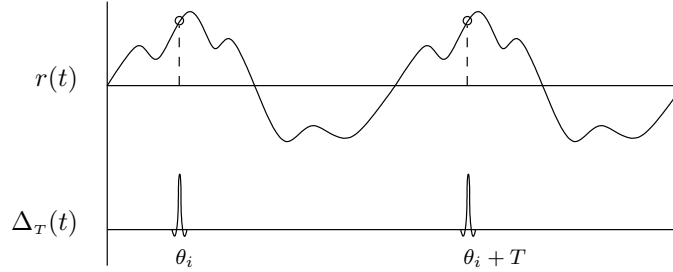
Figure 3.3: Data sampling at a fixed phase. Given a periodic displacement field with fundamental frequency $1/T$ Hz, a snapshot of the tissue can be measured at time θ_i by sampling only at integer multiples of the period. In the above figure, r is a component of the displacement field evaluated at a single point, and Δ_T is a pulse train.

In order to overcome the limitations of the scanner, it must be assumed that the tissue's response reaches a periodic quasi-steady state with a known fundamental frequency, $\omega_o = 2\pi/T$ rad/s. To achieve this, a sinusoidal mechanical excitation is applied, and measurements are started only after a sufficient amount of time has passed to allow for the decay of all transient behaviour. A 'snapshot' can be obtained by sampling at a frequency equal to the driving frequency. After every T seconds, the field returns to its initial configuration, allowing for more measurements to be taken. This is depicted in Figure 3.3.

Suppose the full displacement field corresponding to a particular phase offset, θ_i, is determined. Since the system is assumed to be time-periodic, θ_i can be interpreted as an instantaneous point in time. If the displacements are then determined at a sequence of offsets that cover a full period of motion, a complete description of the time-dependent motion can be extrapolated.

3.2.2 Physics of MR

In order to understand the encoding methods of MRI, the physics of magnetic resonance is reviewed. Quantum mechanics is only briefly mentioned in order to support the concept of an average magnetic moment. After this point, the MRI experiment is described using classical mechanics.

Spins

Magnetic resonance imaging takes advantage of hydrogen nuclei in the body by exploiting a property inherent to all elementary particles known as spin. These protons behave as if they are rotating about some central axis. This has two implications: they have an angular momentum; and since protons are charged particles, they have a magnetic moment.

In the presence of an external magnetic field, the axis of rotation of these particles will align in one of several possible orientations, each possessing an associated *spin number*, s. Protons have two possible values for the spin number: $s = \pm 1/2$. This means that a single proton (or *spin*) can only exist in one of two possible states: *spin-up*, where the magnetic moment is aligned with the same orientation as magnetic field, or *spin-down*, where the moment aligns in the opposite direction. This is depicted in Figure 3.4. According to quantum mechanics, each proton is constantly jumping between the two states. However, the spin-up state is a lower energy state, so is preferred by the proton. Thus, one is more likely to find particles in the 'up' state. When observing a sufficiently large collection of spins, N, the ratio of the number of spins in the upper energy state, N_-, to the lower energy state, N_+, is roughly constant:

$$\frac{N_-}{N_+} = e^{-\Delta E / kT},$$

where ΔE is the energy difference between the two states, $k = 1.38 \times 10^{-23}$ J/K, and T is the temperature. A group, or *packet*, of spins will therefore have a stable average magnetic moment. This allows the magnetic resonance experiment to be described at a macroscopic scale, where a tissue is composed of spin packets, each with a net magnetic moment, $\boldsymbol{M}$, that is aligned in the direction of the magnetic field, and a net angular momentum, $\boldsymbol{I}$. These two quantities are related through the *gyromagnetic ratio*,

$$\boldsymbol{M} = \gamma \boldsymbol{I}.$$

The gyromagnetic ratio for protons (hydrogen nuclei) is 42.58 MHz/T. Each spin packet is thus modelled as a magnetic dipole with an angular momentum.

(a) No background magnetic field   **(b)** Strong background magnetic field
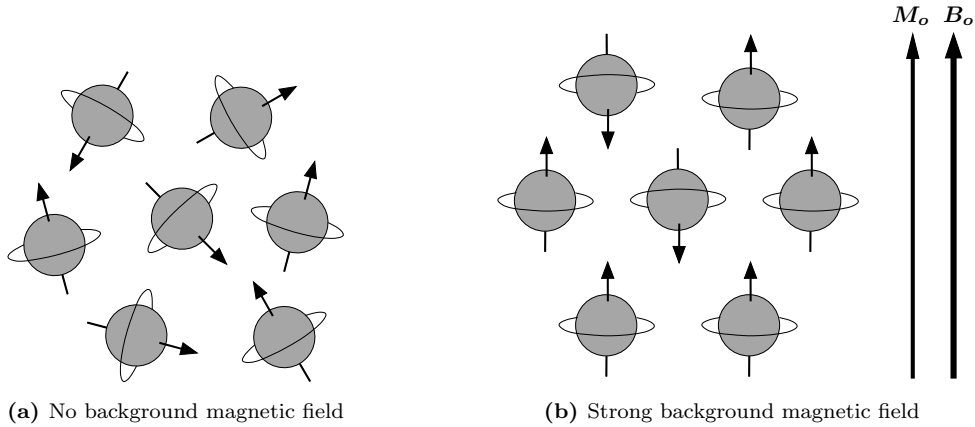
Figure 3.4: In the presence of a magnetic field, $\boldsymbol{B_o}$, the protons will align in one of two possible orientations, generating a net magnetization vector, $\boldsymbol{M_o}$.

The Larmor Frequency

If a magnetic moment is not aligned with the magnetic field, it will precess, much like a spinning top precesses once its axis of rotation becomes misaligned with gravity. At a particular instant in time, the

dipole can be described by the angle it makes with the magnetic field, α, and the phase of its precession, β. These angles are labeled in Figure 3.5. If $\alpha > 0$, the background magnetic field will cause a torque that will change the angular momentum according to

$$\frac{d\boldsymbol{I}}{dt} = \boldsymbol{M} \times \boldsymbol{B_o}.$$

Multiplying both sides by the gyromagnetic ratio and using the relationship between magnetic moment and angular momentum, the following equation is obtained:

$$\frac{d\boldsymbol{M}}{dt} = \gamma \boldsymbol{M} \times \boldsymbol{B_o}.$$

This describes a precession of $\boldsymbol{M}$ about $\boldsymbol{B_0}$, with a frequency $\boldsymbol{\omega} = -\gamma \boldsymbol{B_o}$. This frequency is known as the *Larmor frequency*. If the magnetic field varies in space, the spin packets in regions of different magnetizations will undergo precession at different frequencies. This is the basis for spatial encoding.
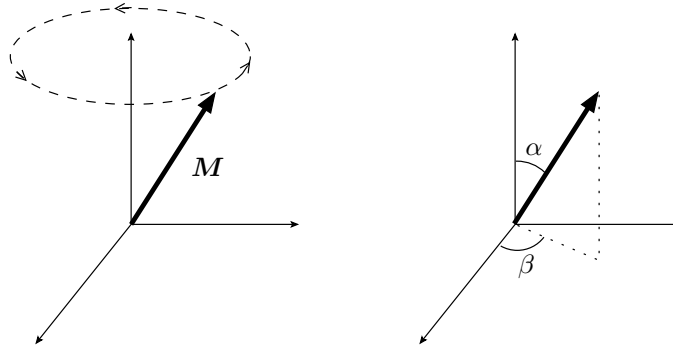
Figure 3.5: If a magnetic moment is not aligned with the background magnetic field, it will undergo precession at the Larmor frequency. At a fixed point in time, the moment can be described by two angles: α, the angle with the background field, and β the precession angle.

The magnetic moment of a spin packet is the sum of the magnetic moments for all the spin dipoles. In the absence of a magnetic field, the orientations of the dipoles are random, so for a sufficiently large number of them, the net magnetic moment is zero. In the presence of a constant magnetic field applied in the z-direction, $\boldsymbol{B_o} = B_o\hat{\boldsymbol{z}}$, the dipoles will begin to precess, slowly aligning with the magnetic field. This creates a net static magnetic moment, $\boldsymbol{M_o} = M_o\hat{\boldsymbol{z}}$. Individual precessing dipoles may have a transverse magnetic moment (moment in the xy-plane) if they are not aligned with the magnetic field. However, since the dipoles have random phases, the transverse components will cancel each other out, resulting in a zero net transverse magnetic moment.

In the MR scanner, the receiving coils are aligned perpendicular to the transverse axis, as shown in Figure 3.6. The transverse moments of the spins rotate, causing a change of magnetic flux through the receiving coils. This generates measurable currents in the coils, according to Faraday's law of induction. These receiver currents are what is measured by the scanner. If a spin packet has no net transverse moment, it will not contribute to the received signal. Therefore, in order to obtain any information, a mechanism is required to tilt the moments away from the z-axis. These transverse moments must also be aligned in phase, or else they will cancel each other out on average, resulting in no net transverse
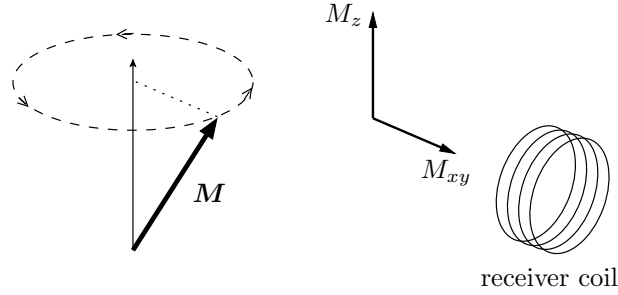
Figure 3.6: MR receiver coil. The receiver coils are aligned to detect only the transverse magnetic moments. Rotating transverse moments induce measurable currents in the coils according to Faraday's law of induction.

moment. This tilting and aligning is accomplished with the use of a radio-frequency (RF) pulse.

RF Pulses

RF coils about the x and y axes of the scanner allow for directed magnetic fields in the transverse plane. If an alternating current is sent through these coils, an alternating magnetic field is produced. By enforcing a 90º phase shift in the current between a coil about the x-axis and one about the y-axis, a rotating magnetic field is generated:

$$\boldsymbol{B_1}(t) = B_1(\sin\omega_1 t + \cos\omega_1 t)$$

All spin packets that have a Larmor frequency ω_1 will experience a constant magnetic field in a rotating frame of reference. They will absorb energy from the RF pulse, and the magnetic moments will tilt away from the z-axis with a common precession phase equal to that of the pulse. This causes a net transverse magnetization that can be picked up by the receiver coils.

The equations of motion for a magnetic moment in a rotating frame, (x', y', z'), with frequency ω_r, is the following:

$$\frac{d\boldsymbol{M}}{dt} = \gamma\boldsymbol{M}\times\boldsymbol{B} - \omega_r\times\boldsymbol{M}$$

$$= \gamma\boldsymbol{M}\times\left(\boldsymbol{B} + \frac{\omega_r}{\gamma}\right),$$

where $\boldsymbol{B}$ is the total magnetic field. For $\boldsymbol{B} = \boldsymbol{B_o} + \boldsymbol{B_1}$, and taking B_1 to be along the x' axis in the rotating frame, this equation is expanded, resulting in

$$\frac{d\boldsymbol{M}}{dt} = \gamma\boldsymbol{M}\times\left[B_1\hat{\boldsymbol{x}}' + \left(B_o + \frac{\omega_1}{\gamma}\right)\hat{\boldsymbol{z}}'\right].$$

Therefore, the angle of the moment, α, will tilt away from the z-axis with angular velocity γB_1. If the

RF pulse is applied for a total time τ, the spin packets will tilt to an angle:

$$\alpha = \gamma B_1 \tau.$$

This is known as the *flip angle*. At this time, the magnitude of the transverse component is

$$M_{xy} = M_o \sin(\gamma B_1 \tau).$$

If the pulse is applied for a time $\tau = 2/\pi \gamma B_1$, a flip angle of $90°$ is created, maximizing the magnitude of the net transverse magnetization. This allows the scanner to read a strong signal. Over time, the tilt angle will realign with the z-axis, and the precession phases of dipoles within each spin packet will de-phase due to dipole-dipole interactions. Both of these processes reduce the amplitude of the net transverse magnetization. This process is known as relaxation.

Relaxation

When the spin packets begin to align with the stationary magnetic field, their projections onto the transverse plane decrease, decreasing the net transverse magnetic field. Meanwhile, the net magnetization in the z-direction will slowly increase until it reaches the static net magnetization, M_o, from before the application of the RF pulse. This is known as *spin-lattice relaxation*, or longitudinal recovery, and occurs at a rate

$$M_z(t) = (1 - e^{-t/T_1})M_o.$$

The time T_1 is called the *spin-lattice relaxation time*.

The dipole-dipole interactions also contribute to a decrease in the net transverse magnetization. These interactions cause the spins within each packet to dephase. The individual transverse moments then begin to average each other out, reducing the net transverse moment of the packet. This relaxation is known as *spin-spin relaxation*, or transverse recovery, and has a corresponding spin-spin relaxation time, T_2:

$$M_{xy}(t) = M_{xy}(0)\, e^{-t/T_2} e^{i\omega_1 t},$$

where $M_{xy}(0)$ is the transverse magnetic moment just after RF excitation, and the real and imaginary components correspond to moments in the x and y directions, respectively. Dephasing can also occur due to inhomogeneities in the magnetic field. This creates a new 'effective' transverse recovery time,

$$T_2^* = \left(\frac{1}{T_2} + \frac{1}{T_{\text{inhomog}}} \right)^{-1}.$$

For purposes of further calculations, this will be assumed to be equal to T_2. The effects of the spin-spin relaxation can be minimized by using spin-echo and gradient-echo sequences. For more information on sequence design, see Haacke *et al.* [22]. The T_2 recovery time is always shorter than the T_1 recovery time. In soft tissues, T_1 is generally on the order of hundreds of milliseconds, whereas T_2 is on the order of tens of milliseconds.

Phase Accumulation

All magnetic moments will also accumulate a phase over time as they continue spinning at the Larmor frequency. If the background magnetic field is allowed to vary in space and time after the RF pulse, this phase is given by the integral:

$$\alpha(\boldsymbol{x}, t) = \gamma \int_0^t \boldsymbol{B}(\boldsymbol{x}, \tau) \cdot \hat{\boldsymbol{z}} d\tau.$$

Let the total applied magnetic field be

$$\boldsymbol{B}(\boldsymbol{x}, t) = [B_o + G_s(\boldsymbol{x}, t) + G_f(\boldsymbol{x}, t) + G_p(\boldsymbol{x}, t) + G_m(\boldsymbol{x}, t)]\, \hat{\boldsymbol{z}}, \tag{3.17}$$

where t is the time after the RF pulse. The fields G_s, G_f and G_p are used to encode spatial information in the signal, and the field G_m is used to encode motion information. Each component of the applied magnetic field contributes to the phase of a spin packet. The final phase shift of the magnetization, M_{xy}, relative to the ending phase of the RF signal, is given by the following:

$$\beta(\boldsymbol{x}, t) = \gamma B_o t + \gamma \int_0^t G_s(\boldsymbol{x}, \tau) d\tau + \gamma \int_0^t G_f(\boldsymbol{x}, \tau) d\tau + \gamma \int_0^t G_p(\boldsymbol{x}, \tau) d\tau + \gamma \int_0^t G_m(\boldsymbol{x}, \tau) d\tau$$
$$= \beta_o(t) + \vartheta_s(\boldsymbol{x}, t) + \vartheta_f(\boldsymbol{x}, t) + \vartheta_p(\boldsymbol{x}, t) + \vartheta_m(\boldsymbol{x}, t). \tag{3.18}$$

Therefore, each spin packet contains a phase contribution from the background magnetization, from the three spatial encoding gradients, and from the motion encoding gradient. This is how information is encoded into the signal. Once these separate phase contributions are recovered, the location and motion information from each spin packet can be extracted.

The equations for the longitudinal and transverse magnetizations of a particular spin packet are

$$M_z(t) = M_o(1 - e^{-t/T_1})$$
$$M_{xy}(\boldsymbol{x}, t) = M_0(1 - e^{-t/T_1}) \exp\left\{-\frac{t}{T_2} + i\,\beta(\boldsymbol{x}, t)\right\}.$$

If RF pulses are sent at a sequence at times with an interval T_R, then the longitudinal component may not be able to fully recover before the next pulse is sent. T_R is known as the time of repetition. This can be included in the magnetization equations:

$$M_z(0^-) = M_o(1 - e^{-T_R/T_1})$$
$$M_{xy}(\boldsymbol{x}, t) = M_o(1 - e^{-T_R/T_1}) \exp\left\{-\frac{t}{T_2} + i\,\beta(\boldsymbol{x}, t)\right\}$$
$$= M_{xy}(\boldsymbol{x}, 0) \exp\left\{-\frac{t}{T_2} + i\,\beta(\boldsymbol{x}, t)\right\},$$

where t is measured as time after the RF pulse.

The transverse moment is the only moment of interest, since it is the one that generates a current in the receiver coils. The time between the RF pulse and the signal recording is known as the *time of echo*, T_E. After this time, it is assumed that the time between successive samples is much smaller than

T_E. Therefore, at the time of measurement, the contribution to the total transverse magnetic field from a spin packet located at position $\boldsymbol{x}$ is given by

$$M_{xy}(\boldsymbol{x}, t) = M_{xy}(\boldsymbol{x}, 0) \, \exp \left\{ -\frac{T_E}{T_2} + i \, \beta(\boldsymbol{x}, t) \right\}$$
$$= \widehat{M}(\boldsymbol{x}, 0) \, \exp \left\{ i \, \beta(\boldsymbol{x}, t) \right\}.$$

The received signal is a sum of all these transverse magnetic moments. The purpose of MRI is to vary the magnetic field in space and time so that information about the individual spin packets can be recovered.

3.2.3   Spatial Encoding

The total signal seen by the receiver coils is a sum of all the transverse magnetic moments:

$$S(t) = \iiint\limits_{V} \widehat{M}(\boldsymbol{x}, 0) \, \exp \left\{ i \, \beta(\boldsymbol{x}, t) \right\} d\boldsymbol{x}.$$

If the phase, β, is constructed such that it is a linear function of time:,

$$\beta(\boldsymbol{x}, t) = a(\boldsymbol{x}) \, t + b(\boldsymbol{x}),$$

then S is simply an inverse Fourier transform, and the transverse magnetic moments can be recovered by taking the Fourier transform of the received signal. The task of spatial encoding is then to linearly incorporate a spin packet's position into this angle, β.

There are two standard techniques for imaging a three-dimensional volume. In the first method, the full volume of interest is excited with an RF pulse. One dimension is encoded in the frequency of the received signal, $a(\boldsymbol{x})$, and the other two dimensions are encoded in the phase, $b(\boldsymbol{x})$. In the second method, the full image is constructed by scanning a sequence of 2D slices of the volume. An RF pulse excites one thin slice of the volume at a time. Each slice is then encoded with one dimension in the frequency of the signal and the other in the phase.

For the MRE application, displacements of the material must also be encoded into the signal. For this reason, multi-slice 2D imaging is often used, with motions encoded in the phase, $b(\boldsymbol{x})$.

Slice Selection

Slice selection is the process of exciting a single slice of the material with an RF pulse. To accomplish this, a magnetic gradient, G_s, is applied during RF excitation, usually along the z-direction:

$$\boldsymbol{B} = \left(B_o + G_s z \right) \hat{\boldsymbol{z}}.$$

The RF pulse is then tuned to the Larmor frequency corresponding to a particular z-value. Only the spins at a height given by

$$z = \frac{B_o}{G_z} - \frac{\omega_1}{\gamma G_z}$$

35

will have a non-zero transverse magnetization. The gradient is then turned off so that all spin packets resume spinning with the base Larmor frequency, $-\gamma B_o$. The tilted spins now have an initial precession phase, ϑ_s.

The RF pulse effectively selects a slice of the volume that will contribute to the received signal. This reduces the problem of reconstruction to two dimensions. These dimensions can then be encoded into the phase and frequency of the signal.

Phase Encoding

After slice selection, all spin packets with a net transverse magnetization begin spinning with the same phase angle. If another linear gradient is applied along the y-axis, then the Larmor frequency will vary linearly along the y-axis, and spin packets will accumulate a phase related to their y-position. After a time τ, the phase accumulation of a spin packet at position (x, y) is given by

$$\vartheta_p(x, y) = -\gamma G_p \tau\, y.$$

When the gradient is turned off, the spin packets will swiftly resume spinning at the base Larmor frequency, but the phase shifts will remain. In this way, the y-coordinates are encoded into the phase of the received signal. All that is further required is to encode the x-coordinates, and devise some method to extract this information.

Frequency Encoding

If a third gradient is applied while the signal is being read, then the recorded signal will contain contributions from spin packets with different Larmor frequencies. By imposing a linear gradient along the x-direction, the spin-packets will gain a precession angle

$$\vartheta_f(x, y) = -\gamma G_f t\, x.$$

In this way, the x-coordinate is encoded into the frequencies of the received signal.

Performing the three spatial encoding gradients in sequence, the corresponding total phase accumulation of a spin packet located at (x, y, z) at time t is given by

$$\begin{aligned}
\beta(x, y) &= \beta_o + \vartheta_s + \vartheta_p + \vartheta_f \\
&= \beta_o + \vartheta_s - \gamma G_p \tau\, y - \gamma G_f t\, x.
\end{aligned}$$

The total received signal, measured by the induced currents in the receiver coils, is thus composed of a sum of frequencies and phases, which reflect the spatial coordinates of all the spin packets.

3.2.4   Motion Encoding

The data required for MRE are measurements of a displacement field. So far, magnetic gradients have been applied in order to encode the spatial information of spin packets. It should come as no surprise that further magnetic gradients are used to encode motion information.

Consider a moving spin packet with an initial position, $\boldsymbol{a}$, and displacement function, $\boldsymbol{r}$. Its position in time in Lagrangian coordinates is

$$\boldsymbol{\varphi}(\boldsymbol{a}, t) = \boldsymbol{a} + \boldsymbol{r}(\boldsymbol{a}, t).$$

If this packet is moving in a background magnetic field, $\boldsymbol{G_m}(\boldsymbol{a}, t)$, then the phase shift is given by

$$\vartheta_m(\boldsymbol{a}, t) = -\gamma \int_0^t \boldsymbol{G_m}(\boldsymbol{a}, \tau) \cdot \boldsymbol{\varphi}(\boldsymbol{a}, \tau) \, d\tau$$

$$= -\gamma \int_0^t \boldsymbol{G_m}(\boldsymbol{a}, t) \cdot \boldsymbol{a} \, dt - \gamma \int_0^t \boldsymbol{G_m}(\boldsymbol{a}, \tau) \cdot \boldsymbol{r}(\boldsymbol{a}, \tau) \, d\tau.$$

If the gradient, $\boldsymbol{G_m}$, is chosen to be periodic with a zero mean, and if it is only applied over a full period, T, then the first integral is zero. Thus, the phase shift at position $\boldsymbol{a}$ is only related to the displacement at $\boldsymbol{a}$:

$$\vartheta_m = -\gamma \int_o^T \boldsymbol{G_m}(\boldsymbol{a}, \tau) \cdot \boldsymbol{r}(\boldsymbol{a}, \tau) \, dt.$$

For this reason, these fields are referred to in the literature as *motion-sensitizing gradients*. Since the displacements in the MRE experiment tend to be small (on the order of a micrometre), the resulting phase shifts due to the motion gradient are small. In order to increase sensitivity, the gradient can be applied for N periods,

$$\vartheta_m = -\gamma \int_0^{NT} \boldsymbol{G_m}(\boldsymbol{a}, \tau) \cdot \boldsymbol{r}(\boldsymbol{a}, \tau) \, dt \tag{3.19}$$

$$= -\gamma N \int_0^T \boldsymbol{G_m}(\boldsymbol{a}, \tau) \cdot \boldsymbol{r}(\boldsymbol{a}, \tau) \, dt. \tag{3.20}$$

This expression is a cross-correlation integral, and can be seen as the projection of the displacement functions onto basis gradients, $\boldsymbol{G_{mi}}$. Therefore, in order to fully recover the displacement functions, a suitable set of basis gradients must be selected.

The motion-sensitizing gradients presented in the original MRE paper [46] are simple square waves, such as the one presented in Figure 3.7. Suppose one wishes to obtain the displacements along the x-direction. Assume the displacement function has the form:

$$r_x(t) = a \cos\left(\omega t - \omega b\right),$$

with unknown amplitude, a, and phase-shift, b. The following motion-sensitizing gradient is applied:

$$\boldsymbol{G_m} = G_m \, f_N \left(\omega t - \omega \theta_i\right) \hat{\boldsymbol{x}},$$

where f_N is a truncated square wave with unit amplitude and N periods. The phase accumulation of
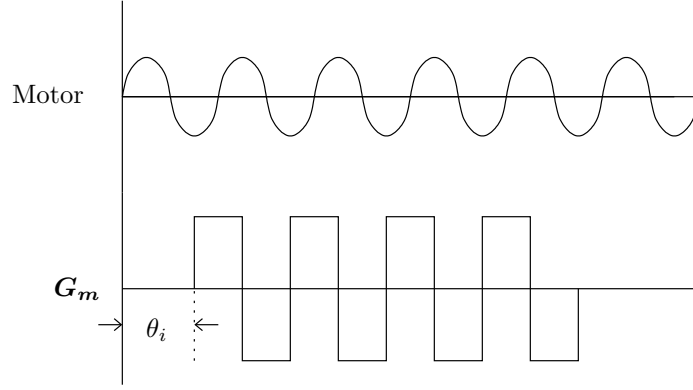
Figure 3.7: A motion-sensitizing gradient, $\boldsymbol{G_m}$, is applied with a frequency equal to the motor's driving frequency. The time offset, θ_i, is varied between scans in order to capture motions at different points in time.

this moving spin packet, given by Equation (3.20), is

$$
\vartheta_m(\theta_i) = -a\gamma N G_m \left[\int_{\theta_i}^{\pi/\omega+\theta_i} \cos(\omega t - \omega b)\,dt + \int_{\pi/\omega+\theta_i}^{2\pi/\omega+\theta_i} \cos(\omega t - \omega b)\,dt \right]
$$

$$
= \frac{4N\gamma\,G_m}{\omega}\,a\sin(\omega\theta_i - \omega b).
$$

In order to recover the parameters a and b, a second scan is required with a different offset, θ_2. Once two measurements are taken, both parameters can be determined uniquely:

$$
\begin{bmatrix} a\sin(\omega b) \\ a\cos(\omega b) \end{bmatrix} = -\frac{\omega}{4N\gamma G_m} \begin{bmatrix} \cos(\omega\theta_1) & -\sin(\omega\theta_1) \\ \cos(\omega\theta_2) & -\sin(\omega\theta_2) \end{bmatrix}^{-1} \begin{bmatrix} \vartheta_m(\theta_1) \\ \vartheta_m(\theta_2) \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}
$$

$$
\implies \quad a = \sqrt{c_1^2 + c_2^2}, \quad b = \left\{ \frac{\arcsin(c_1/a)}{\omega} \right\} \cap \left\{ \frac{\arccos(c_2/a)}{\omega} \right\}, \tag{3.21}
$$

where the inverse trigonometric functions are defined having the range $[0, 2\pi)$. Thus, if the components of the displacement function are sinusoids at the driving frequency, the full displacement field can be recovered with six full scans: two offsets for each of the three dimensions.

The square-wave motion-sensitizing gradient is useful if the displacements are sinusoids. The only restriction placed on displacements in Section 3.2.1, however, is that they are periodic. This means they may contain higher harmonic frequencies. The effect of these harmonics can be explored by considering the displacement component:

$$
r_x^k(t) = a\cos\big(k\omega(t - b)\big).
$$

Repeating the previous calculations, the phase accumulation given by Equation (3.20) due to harmonic

terms is the following:

$$\vartheta_m^k(\theta_i) = \frac{2N\gamma\,G_m}{k\,\omega}\,a\bigg(\sin\big(k\pi - k\omega(\theta_i - b)\big) + \sin\big(k\omega(\theta_i - b)\big)\bigg)$$

$$= \begin{cases} 0, & k \text{ even} \\[2mm] \dfrac{4N\gamma\,G_m/\omega\,\,a\sin\big(k\omega(\theta_i - b)\big)}{k}, & k \text{ odd} \end{cases}$$

Thus, the square wave gradient eliminates even harmonics, and attenuates odd harmonics by a factor of k^{-1}.

Harmonics arise due to non-linearities, particularly at interfaces between materials. The continuum model of the tissue assumes there are no interfaces in the region of interest; however, harmonics may be generated outside the region and reflected back. Regardless, it is expected that any higher harmonics present in the displacement field have much smaller amplitudes than displacements at the fundamental frequency. Since the square-wave motion-sensitizing gradient further attenuates higher frequencies, the displacement function found by Equation (3.21) may be an excellent approximation to the true first-harmonic of the displacement.

In order to avoid making this approximation of ignoring harmonics, a sinusoidal motion-sensitizing gradient can be used:

$$\boldsymbol{G_m}(t) = G_m \cos(\omega t - \omega \theta_i)\,\hat{\boldsymbol{x}}.$$

Since this gradient is applied over an integer multiple of periods, it is easily shown that all higher harmonic components of the displacement are eliminated. Furthermore, if the two offsets are chosen such that they differ by a quarter of a period, the measured phase accumulations are closely related to the real and imaginary components of the first Fourier series coefficient:

$$\mathcal{F}(r_x, \omega) = \frac{\vartheta_m\,(\theta) - i\,\vartheta_m\,\big(\theta + \frac{\pi}{2\omega}\big)}{-\gamma N G_m}, \tag{3.22}$$

where $\mathcal{F}$ is used to denote the Fourier transformation in time. Thus, a sinusoidal gradient is clearly favourable over the square-wave gradient. However, similar to the square-wave result, only the first harmonic of the displacement is obtained with two measurements.

To retrieve all harmonic components of the displacement, a new approach is required. By re-writing Equation (3.20) in a particular way, the phase accumulation due to motion, ϑ_m, can be reinterpreted as the output of a linear filter [51]:

$$\vartheta_m(\boldsymbol{a}, \theta_i) = -\gamma \int_{-\infty}^{\infty} W(\tau)\,\boldsymbol{G_m}(\boldsymbol{a}, \tau) \cdot \boldsymbol{r}(\boldsymbol{a}, \tau)\,dt,$$

where W is a data window (boxcar function) with unit amplitude and support over the interval $[\theta_i, \theta_i + NT]$. Assume the motion-sensitizing gradient has only one non-zero component,

$$\boldsymbol{G_m}(\boldsymbol{a}, \tau) = G_m(\boldsymbol{a}, \tau)\,\hat{\boldsymbol{x}}.$$

If a new function, h, is defined such that $h(t) = -\gamma W(\theta_i - t) G_m(\theta_i - t)$, then the phase accumulation of a spin packet initially at point $\boldsymbol{a}$ can be written as a convolution:

$$\vartheta_m(\theta_i) = \int_{-\infty}^{\infty} h(\theta_i - \tau) \, r_x(\tau) \, d\tau$$

$$= h(\theta_i) * r_x(\theta_i).$$

As described in Section 3.2.1, the phase offsets, θ_i, can be interpreted as a sequence of points in time. Therefore, the discrete Fourier transform can be applied:

$$\Theta(n\omega) = H(n\omega) \, R_x(n\omega),$$

where capital letters denote the Fourier transformed data. This expression is easily inverted to find the harmonics of the displacement, $R_x(n\omega)$, as long as $H(n\omega)$ is non-zero. For this to hold, G_m must have power at all desired harmonic frequencies.

Once a suitable motion-sensitizing gradient is chosen, the motion information can be then encoded into the phase via Equation 3.19. It is important that the gradient along each dimension and at each offset is applied in a separate scan. Therefore, if there are n different time-offsets at which to obtain displacements, then $3n$ separate images must be captured.

3.2.5   Decoding the MR Signal

When motion encoding is combined with spatial encoding, the total signal measured by the receiver coils at time t is given by the following:

$$S(t) = \iiint_V \widehat{M}(\boldsymbol{x}, 0) \, \exp\left\{i\,\beta(\boldsymbol{x}, t)\right\} d\boldsymbol{x}$$

$$= \iiint_V \widehat{M}(\boldsymbol{x}, 0) \, \exp\left\{i\left[\beta_o(\boldsymbol{x}, t) + \vartheta_s(\boldsymbol{x}) - \gamma G_p \tau \, \boldsymbol{x} \cdot \hat{\boldsymbol{y}} - \gamma G_f t \, \boldsymbol{x} \cdot \hat{\boldsymbol{x}} + \vartheta_m(\boldsymbol{x}, \theta_i)\right]\right\} d\boldsymbol{x}. \qquad (3.23)$$

The phase accumulation due to motion is described in Lagrangian coordinates. However, it is known that the displacements are extremely small, relative to the spacial resolution. For the physical experiments considered in this dissertation, the displacements are on the order of a micrometre, whereas the spatial resolution of the scan is on the order of a millimetre. The variation of the displacements is also assumed small relative to the spatial resolution; otherwise the scan would not be able to capture an accurate description of the motion. These large differences in scale allow for the approximation: $\boldsymbol{r}(\boldsymbol{a}, t) \approx \boldsymbol{r}(\boldsymbol{x}, t)$. Therefore, $\vartheta_m(\boldsymbol{a}, \theta_i) \approx \vartheta_m(\boldsymbol{x}, \theta_i)$.

The integral in Equation (3.23) is further simplified due to slice-selection. The initial transverse magnetization is zero for all spin packets not on the excited slice. This converts the received signal into a two-dimensional integral:

$$S(t) = \iint_{\text{slice}} \widehat{M}(\boldsymbol{x}, 0) \, \exp\left\{i\left[\beta_o(\boldsymbol{x}, t) + \vartheta_s(\boldsymbol{x}) + \vartheta_m(\boldsymbol{x}, \theta_i)\right]\right\} \exp\left\{-i\gamma G_p \tau \, y - i\gamma G_f t \, x\right\} dx \, dy.$$

Define the following terms:

$$k_x(t) = \gamma G_f t, \quad k_y(G_p) = \gamma G_p \tau, \quad \tilde{M}(\boldsymbol{x}, t) = \widehat{M}(\boldsymbol{x}, 0) \exp\left\{i\left[\beta_o(\boldsymbol{x}, t) + \vartheta_s(\boldsymbol{x}) + \vartheta_m(\boldsymbol{x}, \theta_i)\right]\right\}.$$

The received signal can be re-written as a function of the new variables k_x and k_y,

$$S(k_x, k_y) = \iint_{\text{slice}} \tilde{M}(\boldsymbol{x}, k_x) \, e^{-i(k_x x + k_y y)} dx \, dy$$

This is exactly the form of a two-dimensional Fourier Transform. Thus, $\tilde{M}$ can be recovered by taking an inverse Fourier transform, provided S is known over a sufficiently large frequency space. The space (k_x, k_y) is known as *k-space* in the literature.

The MR signal is read along one line of k-space at a time. Let the k-space grid have dimensions $N_x \times N_y$. Also, let ΔG_p by a magnitude by which the phase-encoding gradient will increase between lines in k-space, and $G_{p,\text{min}}$ be the initial magnitude. The complete coverage of k-space can be accomplished with the following algorithm:

1. Initiate variable: j=0
2. Slice encoding:
   ○ Turn on slicing gradient G_s, apply RF pulse to select z_{slice}, then remove G_s.
3. Phase encoding:
   ○ Set k_y by taking $G_p = G_{p,\text{min}} + j\Delta G_p$, then apply the gradient for time τ.
4. Motion encoding:
   ○ Apply $\boldsymbol{G_m}$ with time-offset θ_i for time NT.
5. Frequency encoding:
   ○ Initiate k_x by applying $G_f = \pm|G_f|$, depending if j is even or odd.
   ○ Sample $S(t)$ at times Δt, covering all N_x points along k_x.
6. Set $j = j + 1$. If $j < N_y$, go to step 2.

The trajectory of this process through k-space is depicted in Figure 3.8. The pulse-sequence diagram for one read-line is shown in Figure 3.9. Once a full image is captured, the algorithm must be repeated for each motion-sensitizing basis function along each spatial dimension, and at each time-offset, θ_i. After these multiple scans, the full displacement field can be recovered for one slice of the three-dimensional volume. The entire process must then be repeated for each slice.

Once all values of $S(k_x, k_y)$ are obtained, an Inverse Discrete Fourier Transform (IDFT) is applied, recovering the values $\tilde{M}(\boldsymbol{x}, k_x)$ at each discrete grid point on the slice. This is not the desired quantity, however; the displacements still need to be extracted. The displacements are encoded in the phase ϑ_m. In order to isolate this term, the entire scanning process is repeated, reversing the polarity of the motion-encoding gradient. By doing so, all phase contributions to $\tilde{M}$ unrelated to motion-encoding will
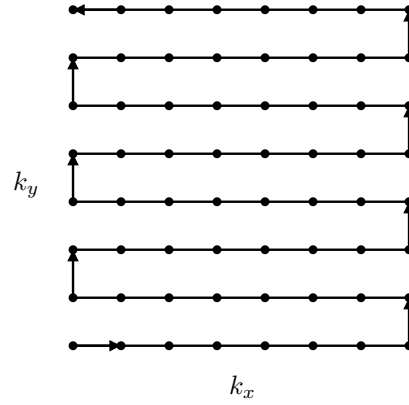
Figure 3.8: During an MR scan, k-space is covered by reading one line along k_x at a time. k_x is increased by sampling in time, and k_y is increased by changing the value of G_p.
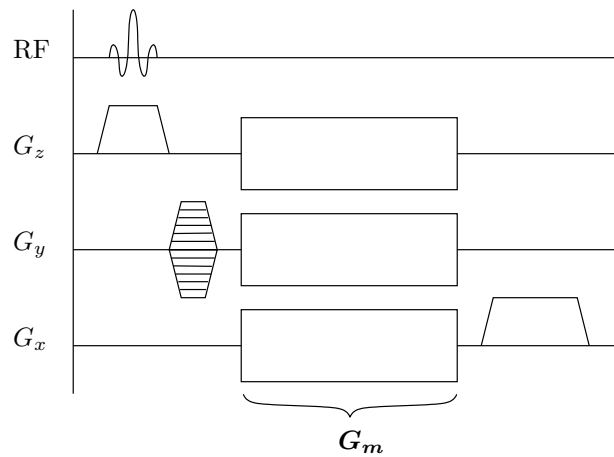
Figure 3.9: Pulse sequence diagram for motion-encoding. The slice-selection gradient is applied along the z-axis, phase-encoding along the y-axis, and frequency-encoding along the x-axis. The motion encoding must be performed along each of the axes in separate scans.

be the same, but the phase-offset due to motion-encoding will be reversed:

$$\tilde{M}_1(\boldsymbol{x}, k_x) = \widehat{M}(\boldsymbol{x}, 0) \, \exp\left\{i\left[\beta_o(\boldsymbol{x}, k_x \,/\, \gamma G_f) + \vartheta_s(\boldsymbol{x}) + \vartheta_m(\boldsymbol{x}, \theta_i)\right]\right\}$$

$$\tilde{M}_2(\boldsymbol{x}, k_x) = \widehat{M}(\boldsymbol{x}, 0) \, \exp\left\{i\left[\beta_o(\boldsymbol{x}, k_x \,/\, \gamma G_f) + \vartheta_s(\boldsymbol{x}) - \vartheta_m(\boldsymbol{x}, \theta_i)\right]\right\}$$

By taking the difference between the two phases (known as the *phase-difference* technique), the phase-accumulation due to motion can be recovered:

$$\angle\widetilde{M}_1 - \angle\widetilde{M}_2 = 2\,\vartheta_m(\boldsymbol{x}, \theta_i).$$

Phase-differencing also removes phase accumulations due to magnetic field inhomogeneities. At this stage, phase-unwrapping techniques might be required in order to ensure continuity. Once the phase ϑ_m is determined at each grid point and time-offset, the techniques of Section 3.2.4 can be used to determine the components of the displacement field.

Instead of repeating the scanning procedure with reversed motion-encoding gradients, the scan can be repeated with no motion encoding. The phase-difference technique is still able to isolate the phase due to motion, although is only half as sensitive. This method requires fewer scans, since there is no need to repeat the second scan for all motion-encoding basis functions. It also produces the standard magnitude image, which can be used for other purposes, such as image registration.

3.2.5.1 Summary

Magnetic resonance imaging allows for high-resolution measurements of the displacement field over a three-dimensional volume. This is the major advantage over ultrasound techniques. The large amount of available data can improve the accuracy of inversion methods. Images of the displacement field are not captured instantaneously; in order to measure the full time-dependent displacements, it must be assumed that they are periodic with a known fundamental frequency. This allows measurements to be performed in bursts, capturing data only at multiples of the fundamental period. This periodicity requirement is reflected in the mathematical model described in the previous section.

The physics behind MRI is based on the concept of spins and spin packets. These spin packets act like spinning, magnetic tops. They have angular and magnetic moments, which precess when not aligned with the background magnetic field. The speed of precession is given by the Larmor frequency, which depends on the strength of the field. When precessing, the spin-packets have non-zero transverse magnetic moments, which contribute to the received signal by inducing currents in receiver coils. Spatial encoding is accomplished by exciting a slice of the volume with an RF pulse, encoding one spatial dimension into the phase of the transverse moments, and encoding the other spatial dimension into the angular frequency of the transverse moments. This is accomplished by applying magnetic gradients during the encoding stages. Motion information is also encoded into the phase of the moments, by applying time-varying motion-sensitizing gradients during the motion-encoding stage. The spatial information in the received signal is decoded by applying an IDFT, and the motion information is obtained by taking the difference in phase between two images. The final result is a full three-dimensional displacement field.

Only a basic pulse sequence was described in this section. More advanced techniques, such as spin echoes and gradient echoes can be used to reduce T_2 decay, allowing for more data to be captured

between RF pulses. There are also techniques to reduce ghosting caused by motion. Descriptions of these methods can be found in Haacke *et al.* [22].

Now that an appropriate mathematical model is derived, and measurements of the full displacement field are obtained, one can attempt the task of estimating the model parameters based on the data. This is a type of inverse problem. Inverse problems are typically ill-posed, which require regularization techniques in order to solve. Some of these techniques are described in the next section.

3.3   Inverse Problems

The tissue model presented in this chapter was derived to describe motions in a material, given a set of parameters. With a complete description of a material, one can use the model to predict the displacement field. This predicted field can be measured, and the model can be verified. The parameters of the system *cause* a certain displacement field, which is a measurable *effect*. This is known as a *forward problem*: given the set of causes, predict the verifiable effect. An *inverse problem* is the opposite: by observing the effects of a system, infer something about the causes. Parameter estimation is a prime example of this type of problem.

Inverse problems are often *ill-posed*. In the sense of Hadamard [23], a well-posed problem satisfies the three following conditions:

· A solution exists.

· The solution is unique.

· The solution depends continuously on the data.

An ill-posed problem is one that does not satisfy at least one of these requirements. When continuous problems are discretized, the third condition no longer formally applies. Its purpose is to eliminate cases where small perturbations of the data drastically alter the solution. When this occurs in the discrete case, the problem is said to be *ill-conditioned*.

3.3.1   Existence

Physical systems are very complex, and can depend on an uncountable set of variables. In order to gain understanding, simple mathematical models are presented that attempt to approximate systems with as few parameters as possible. There is a very famous statement by the statistician, George Box:

"All models are wrong, some models are useful [9]."

Models are not exact, and cannot include every factor that influences a system. Because of this, there is no perfect parameter that can describe every single observed output; there can only be parameters that best approximate, or *best fit* the data. This idea is best described by a simple parameter identification example.

Example 1.1. *Determine the parameters a and b in the following model that best fit the provided data:*

$$y = ax + b, \quad \begin{array}{c|cccc} x & 1 & 2 & 3 & 4 \\ \hline y & 2 & 3 & 5 & 3.5 \end{array}$$

This problem has no solution. Using only the first two points, the values of the parameters are found to be $a = 1$, $b = 1$. However, these parameters do not describe the third point, since $5 \neq 3(1) + 1$. Therefore, there is no solution. Even though no perfect solution exists, a useful answer can still be obtained. Converting the problem into matrix form results in the following overdetermined linear system:

$$
\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 3.5 \end{bmatrix}
$$

Since there is no single solution for (a, b), one can instead try to find the best-fitting solution. The above is in the form $Y = X\beta$. The standard technique to solve this is to minimize the squared error in the sense of ℓ_2:

$$
\beta = \arg\min_{\beta} \ \|\beta X - Y\|^2
$$

$$
\implies \quad 0 = \frac{d}{d\beta} \|X\beta - Y\|^2
$$

$$
\implies \quad \beta = (X^{\text{H}}X)^{-1} X^{\text{H}}Y.
$$

This is the *Ordinary Least-Squares* (OLS) solution. In this example, the OLS solution is the following:

$$
\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 37 \\ 13.5 \end{bmatrix}
$$

$$
= \begin{bmatrix} 0.65 \\ 1.75 \end{bmatrix}
$$

Therefore, the model that best fits the data is $y = 0.65x + 1.75$.

Ordinary least-squares is useful in finding a best-fitting solution when no exact solution exists. In the method, all equations are treated as being equally important. If it is known that some of the equations are more reliable than others, a weighted least-squares estimate can be found:

$$
\beta = \arg\min_{\beta} \ \left\| \hat{W} \left(X\beta - Y \right) \right\|^2
$$

$$
\implies \quad \beta = \left(X^{\text{H}} \hat{W}^{\text{H}} \hat{W} X \right)^{-1} X^{\text{H}} \hat{W}^{\text{H}} \hat{W} Y,
$$

where $\hat{W}$ is a matrix of weights. Let $W = \hat{W}^{\text{H}} \hat{W}$ be the weighting matrix, but remember that this is a Hermitian matrix, and that for diagonal matrices, the entries are now squared weights. The weighting matrix, W, can be designed to decorrelate noise and minimize the variance in the estimate. This is the purpose of the *Generalized Least-Squares* (GLS) method. The minimizing W satisfies $W = \Sigma^{-1}$, where Σ is the correlation matrix of the noise in measurements of Y. It is possible to find this W by Cholesky factorization. Consider the following modified example:

Example 1.2. *Determine the best-fitting parameters a and b in the following system:*

$$
\begin{bmatrix} 2 \\ 3 \\ 5 \\ 3.5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} 0.5\,\epsilon_1 \\ 0.5\,\epsilon_2 \\ 1.5\,\epsilon_3 \\ 1.0\,\epsilon_4 \end{bmatrix},
$$

where ϵ is white Gaussian noise with unit variation.

The generalized least-squares method has the following parameters:

$$
\Sigma = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 2.25 & 0 \\ 0 & 0 & 0 & 1.00 \end{bmatrix} \qquad W = \begin{bmatrix} 4.00 & 0 & 0 & 0 \\ 0 & 4.00 & 0 & 0 \\ 0 & 0 & 0.44 & 0 \\ 0 & 0 & 0 & 1.00 \end{bmatrix},
$$

which admits the following best estimate and covariance matrix:

$$
\begin{bmatrix} a \\ b \end{bmatrix} = \left(X^{\mathsf{H}} \Sigma^{-1} X \right)^{-1} X^{\mathsf{H}} \Sigma^{-1} Y \qquad\qquad \mathrm{var}\left\{ \begin{bmatrix} a \\ b \end{bmatrix} \right\} = \left(X^{\mathsf{H}} \Sigma^{-1} X \right)^{-1}
$$

$$
= \begin{bmatrix} 0.667 \\ 1.5 \end{bmatrix} \qquad\qquad = \begin{bmatrix} 0.122 & -0.224 \\ -0.224 & 0.517 \end{bmatrix}
$$

The formula for the covariance matrix is easily derived from the original expression, where ϵ is the only random component, and the identity $\mathrm{var}\{A\epsilon\} = A \, \mathrm{var}\{\epsilon\} A^{\mathsf{H}}$ is used.

In the previous example, the variances of the noise were known. In most practical applications, the noise statistics are not known. For this reason, *Feasible Generalized Least-Squares* (FGLS) was developed. The first step in FGLS is to find the OLS solution. The noise is then assumed to be uncorrelated, and an estimate of the noise variances is taken to be the squared residuals. With this covariance, GLS is employed to find the best solution. This process can be iterated, where the covariance for the next iteration is estimated as follows:

$$
\Sigma_{i+1} = \mathrm{diag}\left\{ (X\beta_i - Y)^2 \right\},
$$

where a subscript i denotes the ith iteration. This can be continued until the system converges, $\Sigma_i \to \widehat{\Sigma}$. The estimated covariance of the parameter, β, is given by:

$$
\mathrm{var}\left\{ \beta \right\} = \left(X^{\mathsf{H}} \Sigma_1^{-1} X \right)^{-1} \left(X^{\mathsf{H}} \Sigma_1^{-1} \widehat{\Sigma} \Sigma_1^{-1} X \right) \left(X^{\mathsf{H}} \Sigma_1^{-1} X \right)^{-1}
$$

The FGLS iterates for Example 1.2 are shown in Table 3.1. The final result is:

$$
\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathrm{var}\left\{ \begin{bmatrix} a \\ b \end{bmatrix} \right\} = \begin{bmatrix} 0.161 & -0.230 \\ -0.230 & 0.328 \end{bmatrix}.
$$

Table 3.1: FGLS iterates for the line of best-fit example.

a	b	σ_1^2	σ_2^2	σ_3^2	σ_4^2
0.65	1.75	0.160	0.002	1.690	0.723
0.71	1.57	0.080	3.2×10^{-5}	1.676	0.840
0.83	1.33	0.028	8.1×10^{-9}	1.362	1.359
$\vdots$			$\vdots$		
1	1	0	0	1	2.25

(a) GLS

(b) FGLS
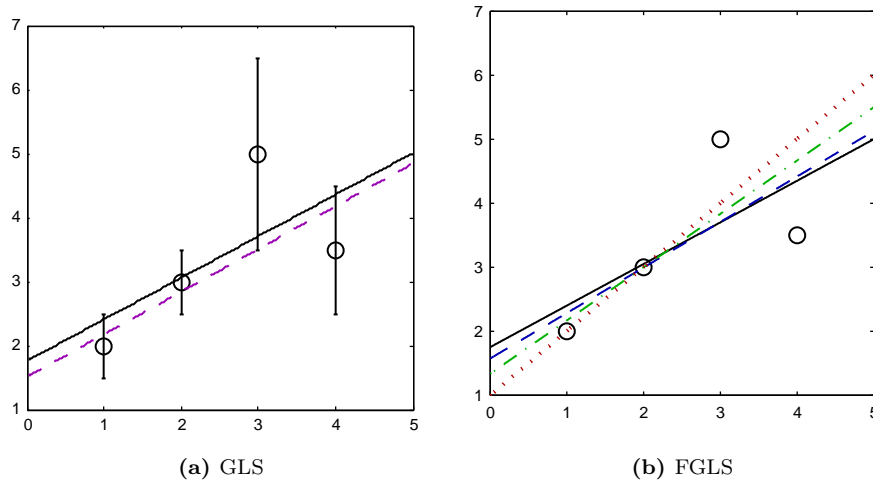
Figure 3.10: Generalized least-squares. Ordinary least-squares (black) minimizes the squared error, assuming outputs contain noise of equal variance. Generalized least squares (purple) minimizes the squared error, given known noise statistics. Feasible generalized least squares (blue → red) can be used when the noise statistics are not known, but assumed to be uncorrelated. It begins with the OLS solution, then estimates noise from the residuals.

The results of the OLS, GLS and FGLS algorithms for Example 1.2 are shown in Figure 3.10. By examining this figure, it can be observed that GLS ensures the model fits closer to the measurements that are known to have less noise than OLS. The FGLS result begins by weighting all measurements equally, but then eventually finds two points to fit the parameters perfectly. To justify ignoring the other two points, large noise variances are estimated for them. This might be unreasonable. The advantage of FGLS is that it can quickly remove the effect of outliers. This disadvantage is that it will eventually isolate very few points on which to fit the parameters. It may be more reasonable to use only one or two iterates. More information on least-squares algorithms can be find in Kariya [27].

With the previous examples, it was shown that even though no formal solution exists, a best-fitting set of parameters could be found. These least-squares algorithms can minimize the effect of noise on the parameter estimates, given some knowledge or assumptions about the noise. This is a type of regularization that helps remove the effect of unknown information, namely noise. Not all unknown information can be treated this way. Sometimes the missing information is key in obtaining the correct solution. Without it, strong assumptions must be made in order to find a best guess of the answer.

3.3.2   Uniqueness

One of the major issues with inverse problems is a lack of information. Not all variations in the causes are directly observable in the effects, making the cause difficult to determine. In order to make an inference, either outside information is needed, or assumptions must be made about the form of the solution.

A typical example of a problem with no unique solution is an underdetermined system of linear equations, such as the following:

$$x + y + z = 3$$
$$y + z = 2.$$

In this case, a general solution can be written explicitly:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad \alpha \in \mathbb{R}$$

It is not always possible to write out a general solution. For example, consider the three-dimensional Laplace equation:

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0.$$

Without specifying some form of extra conditions on f, like boundary conditions, this problem has no unique solution. No general solution can be explicitly written either. In order to infer a specific f, some assumption or condition must be placed on the set of solutions. This idea of introducing additional information is known as *regularization*.

3.3.2.1   Quadratic Programming

The most common strategy for handling non-uniqueness is to select the solution that minimizes some functional. Consider an underdetermined discrete system of the form $Ax = b$. Since the system is underdetermined, there is no unique solution. However, one can incorporate outside information or assumptions about the desired the solution, and re-pose the problem in the following form:

$$\begin{aligned} &\textit{Minimize} \qquad f(x) = x^{\mathrm{H}} Q x + c^{\mathrm{H}} x + x^{\mathrm{H}} c, \\ &\textit{subject to} \qquad Ax = b. \end{aligned}$$

This is known as an *Equality-constrained Quadratic Programming* (EQP) problem. The matrix Q is restricted to the set of Hermitian, positive-semidefinite matrices. With this condition, f is guaranteed to be convex, allowing a minimum to exist. This provides a general framework for solving underdetermined systems: impose the original problem, $Ax = b$, as an equality condition, and select Q and c to steer towards the desired solution. Methods for solving such a problem are described in Nocedal [48]. These can be divided into two classes: *null-space* methods and *range-space* methods.

Null-space methods can always be used in EQP. Let $x \in \mathbb{C}^n$, and rank$\{A\} = m < n$. Assume A has full rank (i.e. has exactly m rows). The first step in these methods is to find a full-rank matrix $Z \in \mathbb{C}^{n \times (n-m)}$, such that range$\{Z\} = $ null$\{A\}$. This is the null-space of A, and is commonly determined with the use of a singular value decomposition [67]. A particular solution, x_p, is required next. One can be found by using the null-space to create a determined system with solution:

$$x_p = \begin{bmatrix} A \\ Z^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

The general solution to the constraint can then be written as

$$x = x_p + Zw, \quad w \in \mathbb{C}^{n-m}.$$

Restricting x to this form guarantees that the equality condition is met. By substituting x back into the quadratic functional, using the symmetry of Q, and removing the constant terms (since they do not affect the minimization), the problem is converted into an unconstrained minimization of

$$\tilde{f}(w) = w^{\mathrm{H}} Z^{\mathrm{H}} Q Z w + (Q x_p + c)^{\mathrm{H}} Z w + w^{\mathrm{H}} Z^{\mathrm{H}} (Q x_p + c).$$

The minimizing w satisfies

$$Z^{\mathrm{H}} Q Z w = -Z^{\mathrm{H}} (Q x_p + c)$$

If $Z^{\mathrm{H}} Q Z$ has full-rank, then this new system has a unique solution:

$$w = - (Z^{\mathrm{H}} Q Z)^{-1} Z^{\mathrm{H}} (Q x_p + c).$$

If $Z^{\mathrm{H}} Q Z$ does not have full rank, then the system is still underdetermined. This means the conditions placed on x are not enough to isolate a particular solution. The steering matrix, Q, needs to be

readjusted. The final solution of the null-space method is

$$x = x_p - Z \left(Z^{\mathrm{H}} Q Z\right)^{-1} Z^{\mathrm{H}} \left(Q x_p + c\right).$$

The procedure was to find a particular solution of the original problem, then add an element of the null-space in order to minimize the quadratic functional.

If Q has full rank, then range-space methods can be used. A simple range-space method can be derived by implementing a Lagrange multiplier:

$$L(x, \lambda) = x^{\mathrm{H}} Q x + c^{\mathrm{H}} x + x^{\mathrm{H}} c + \lambda^{\mathrm{H}} \left(A x - b\right) + \left(A x - b\right)^{\mathrm{H}} \lambda.$$

There are two directions one can proceed. The simplest is to convert this into a larger system,

$$L(x, \lambda) = \begin{bmatrix} x^{\mathrm{H}} & \lambda^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} Q & A^{\mathrm{H}} \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} c^{\mathrm{H}} & -b^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} x^{\mathrm{H}} & \lambda^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} c \\ -b \end{bmatrix},$$

which has the solution:

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} Q & A^{\mathrm{H}} \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} -c \\ b \end{bmatrix}.$$

The desired solution can then be extracted from the above vector. Every time this problem is solved, a solution vector and a Lagrange multiplier are computed. This enlarged problem can be up to twice as big as the original, which may put a strain on computational resources. If the algorithm will be repeated many times for a single regularization matrix, Q, it may be faster to pre-compute the inverse of Q and solve for the Lagrange multiplier explicitly:

$$\begin{aligned}
\frac{1}{2} \frac{\partial}{\partial x} L(x, \lambda) &= Q x + c + A^{\mathrm{H}} \lambda \\
\implies \quad 0 &= x + Q^{-1} c + Q^{-1} A^{\mathrm{H}} \lambda \\
\implies \quad 0 &= b + A Q^{-1} c + A Q^{-1} A^{\mathrm{H}} \lambda \\
\implies \quad \lambda &= - \left(A Q^{-1} A^{\mathrm{H}}\right)^{-1} \left(A Q^{-1} c + b\right).
\end{aligned}$$

With this multiplier, the following minimizing solution is obtained:

$$x = Q^{-1} \left[A^{\mathrm{H}} \left(A Q^{-1} A^{\mathrm{H}}\right)^{-1} \left(A Q^{-1} c + b\right) - c\right].$$

There is no need to compute the null-space of A. This can drastically reduce computation time. Often, Q has a very simple form and is easy to invert. For example, if one wishes to determine the solution that has the smallest ℓ_2-norm, then Q is simply the identity matrix. This makes range-space methods more computationally efficient. The disadvantage is that the form of Q is limited. Null-space methods apply to more general situations.

The inverse problem in MRE is to solve a partial differential equation for the shear modulus, without knowledge of the boundary conditions. Consider the following one-dimensional example.

Example 2.1. *Determine the shear modulus, $\mathcal{M}$, in the one-dimensional equation:*

$$- \sin(x) = \frac{d}{dx}\left(\mathcal{M}(x)\, 2\, \frac{d}{dx}\sin(x)\right), \quad x \in [0, \pi/4]. \tag{3.24}$$

This is the 1D version of Equation (3.15), with $\omega = 1$ and $U(x) = \sin(x)$. Simply integrating results in the following general solution:

$$\mathcal{M}(x) = \frac{1}{2} + \frac{c}{2\cos(x)}, \qquad c \in \mathbb{C}.$$

Suppose one wishes to obtain the solution with the smallest L_2-norm. Then, the task is to determine the free parameter, c, that minimizes

$$f(c) = \int\limits_0^{\pi/4} \left(\frac{1}{2} + \frac{c}{2\cos(x)}\right)^2 dx.$$

Setting the derivative of f to zero and isolating c gives

$$c = -\ln\left(1 + \sqrt{2}\right).$$

Therefore, the solution with the smallest L_2-norm is

$$\mathcal{M}(x) = \frac{1}{2} - \frac{\ln\left(1 + \sqrt{2}\right)}{2\cos(x)}.$$

In this way, a particular solution is isolated that satisfies the ill-posed problem. There is no guarantee that this matches 'true' solution. In order for it to be reasonable, the assumptions used to regularize the problem must be well-founded.

In terms of the elasticity problem, it may not be reasonable to expect the material to have a shear modulus with the smallest norm. In fact, the obtained solution has negative values, which is not admissible from a physical point of view. Perhaps a better assumption is that the variation of the shear modulus is minimal. The norm of the derivative of $\mathcal{M}$ can be minimized with the functional:

$$f_2(c) = \int\limits_0^{\pi/4} \left[\frac{d}{dx}\left(\frac{1}{2} + \frac{c}{2\cos(x)}\right)\right]^2 dx.$$

The minimizing argument is $c = 0$, which results in the constant shear modulus: $\mathcal{M} = 1/2$. This might be more reasonable to expect in a material.

Most practical problems are discretized in order to solve them numerically. In MRE, the data consist of displacement measurements at discrete points on a grid. Therefore, the partial differential equation must be discretized and solved numerically. This is done for the previous continuous example.

Example 2.2. *Determine the solution vector, $\mathcal{M}$, that satisfies the discrete differential equation:*

$$-U = 2D \left(\operatorname{diag}\{\mathcal{M}\} DU\right),$$

where D is a discrete derivative operator, and

$$U = \left[\sin(0),\ \sin\left(\frac{\pi}{4N}\right),\ \sin\left(\frac{2\pi}{4N}\right), \ldots,\ \sin(\pi/4)\right]^T.$$

In order to solve for $\mathcal{M}$, one of the derivative operators must be inverted. Just like integration, this derivative inversion must account for boundary conditions. In this case, however, the boundaries are unknown. The discrete operator, D, is not full rank. In particular, the vector $[1, 1, \ldots, 1]^T$ is in its null-space. In other words, if x is a vector of constants, $Dx = 0$, which is expected from a derivative operator. Therefore, the inversion does not have a unique solution; there is one free parameter that corresponds to a boundary condition.

Let $A = 2D$, $b = -U$, and $x = \operatorname{diag}\{\mathcal{M}\} DU$. The new system is then of the form $Ax = b$, which is underdetermined. In order to obtain a solution, some regularizing conditions must be set in place.

In the continuous case, the first solution found minimized the L_2-norm of the function. The discrete version of this is to minimize the ℓ_2-norm of $\mathcal{M}$,

$$\begin{aligned}
\textit{Minimize} \quad f(\mathcal{M}) &= \mathcal{M}^{\mathrm{H}}\mathcal{M} \\
&= x^{\mathrm{H}} \left(\operatorname{diag}\{DU\}^{-1}\right)^{\mathrm{H}} \operatorname{diag}\{DU\}^{-1} x \\
&= \tilde{f}(x).
\end{aligned}$$

Since $\operatorname{diag}\{DU\}$ is diagonal, its inverse simply contains the reciprocal of the entries along the diagonal. If any of these entries are zero, the Moore-Penrose pseudoinverse is considered. The problem is now posed in the form required by EQP, with the following parameters:

$$Q_1 = \left(\operatorname{diag}\{DU\}^{-1}\right)^{\mathrm{H}} \operatorname{diag}\{DU\}^{-1}, \qquad c = 0.$$

It is easily seen that Q_1 is Hermitian and positive-semidefinite:

$$Q_1^{\mathrm{H}} = \left[\left(\operatorname{diag}\{DU\}^{-1}\right)^{\mathrm{H}} \operatorname{diag}\{DU\}^{-1}\right]^{\mathrm{H}} \qquad\qquad x^{\mathrm{H}}Q_1 x = \left\|\operatorname{diag}\{DU\}^{-1} x\right\|^2$$
$$= Q_1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad \geq 0$$

Therefore, the previously described null-space or range-space methods can be used.

Similarly, the minimum variation regularizing condition on $\mathcal{M}$ can be imposed discretely, with the minimizing functional:

$$f(\mathcal{M}) = \mathcal{M}^{\mathrm{H}} D^{\mathrm{H}} D \mathcal{M}$$
$$\tilde{f}(x) = x^{\mathrm{H}} \underbrace{\left(\operatorname{diag}\{DU\}^{-1}\right)^{\mathrm{H}} D^{\mathrm{H}} D \operatorname{diag}\{DU\}^{-1}}_{=Q_2} x$$

(a) Minimum norm   (b) Minimum variation
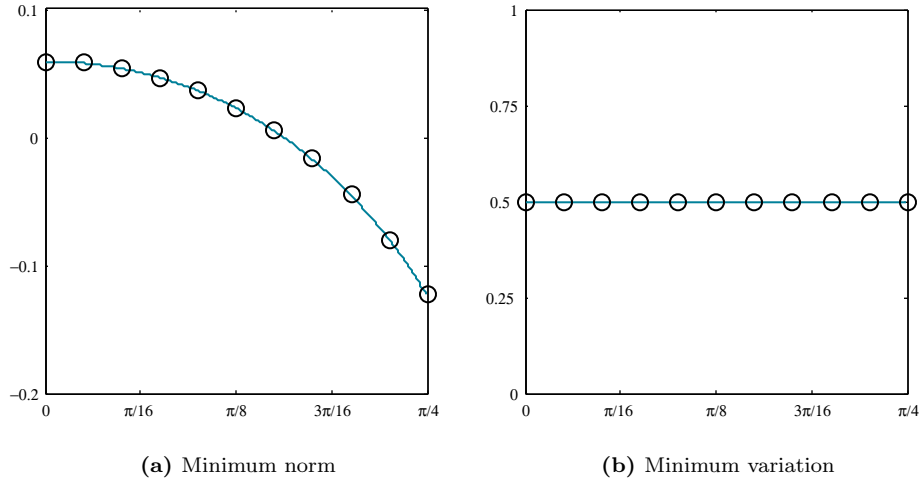
Figure 3.11: Solutions to the one-dimensional elastic wave equation. Both continuous (blue) and discrete (black) versions of Equation (3.24) are solved for the shear modulus, $\mathcal{M}$. Since boundary conditions are unknown, regularizing conditions are set in place to guarantee a unique solution.

Again, it is clear that this Q_2 is Hermitian and positive semidefinite.

It is expected that these discrete versions of the one-dimensional elasticity problem will have the same solutions as the continuous ones derived previously. This is verified in Figure 3.11. For both regularization conditions, the discrete solutions perfectly correspond to the continuous ones. This strongly suggests that equality-constrained quadratic programming will be useful in shear-modulus inversions from the full three-dimensional differential equation, when there is no knowledge of the boundary conditions.

3.3.3   Numerical Conditioning

In the previous section, methods of solving ill-posed problems were discussed. When multiple solutions exist, one can be singled out by applying a regularizing condition. When no formal solution exists, a practical one can be determined in which the error is minimized. There are situations where a unique, formal solution does exist, but is not to be trusted from a practical point of view. Measurements are subject to noise, digital computations introduce round-off errors, and the discrete approximation of continuous functions and operators is subject to approximation errors. For these reasons, one must often question the reliability of a solution. Numerical problems may be very sensitive to small variations. If these small errors and approximations in the data can cause drastically different results, then the problem is said to be *ill-conditioned*. Such problems are usually numerically unstable, and require regularization techniques to stabilize them. A simple linear systems example can help demonstrate this.

Example 3.1. *Determine the unknown variables, x_1 and x_2, in the following system of linear equations:*

$$1.000\, x_1 + 1.000\, x_2 = 2.000 \tag{3.25a}$$
$$1.000\, x_1 + 1.001\, x_2 = 2.001. \tag{3.25b}$$

This problem has the solution:

$$(x_1,\, x_2) = (1,\, 1)$$

Now, suppose that the coefficients were obtained from noisy measurements. Can the precise values be trusted? By slightly perturbing the right-hand side of Equation (3.25b) from 2.001 to 2.002, the formal solution is changed to

$$(x_1,\, x_2) = (0,\, 2)$$

This high sensitivity to variations in the data makes this problem ill-conditioned.

The condition of a numerical problem is often described by a *condition number*. This number expresses the expected amount of change in the solution given a change in the data. For linear systems of the form $Ax = b$, this number is rather easy to define. The condition number is inherent to the matrix A, and is given by

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)},$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and smallest singular values of A, respectively. If the condition number is extremely large, the problem is deemed ill-conditioned. Otherwise, the problem is well-conditioned. The condition number for Example 3.1 is $\kappa = 4000$. This means that varying the right-hand side by 0.001 can change the values of the unknown parameters by a value of four.

From a practical point of view, the two equations presented in the example are equal, so the system is effectively underdetermined. The methods discussed in the previous section will be of no help in this instance; if the existing system is set as an equality constraint, the same formal solution will be determined. A new form of regularization is required.

3.3.3.1 Tikhonov Regularization

The most common regularization method for ill-conditioned problems is Tikhonov regularization. In this method, an extra term is added to the least-squares minimization problem to favour desirable solutions:

$$x = \arg\min_{x} \left\{ \|Ax - b\|^2 + \|\Gamma x\|^2 \right\},$$

where Γ is known as the Tikhonov matrix. This can be generalized by applying generalized least-squares. The new problem has the solution:

$$x = \left(A^{\mathrm{H}}A + \Gamma^{\mathrm{H}}\Gamma\right)^{-1} A^{\mathrm{H}}b.$$

This solution attempts to balance the two terms in the minimization, weighted by the relative norms of A and Γ. The advantage of Tikhonov regularization is that it is incredibly simple to implement. The only difficulty is in selecting an appropriate Γ. If the original matrix problem is well-conditioned, the Tikhonov matrix should have a relatively small norm to avoid steering the solution away from satisfying the original system. If the problem is ill-conditioned, it should have a higher weight in order to overcome the numerical instability.

A typical Tikhonov matrix is $\Gamma = \alpha I$, where α is a weighting factor to optimize the balance between the two terms in the minimization. This α is often chosen in an ad-hoc fashion. If entries in x and b are independent and normally distributed, then it can be shown that the optimal choice is $\alpha = \sigma_b/\sigma_x$. Applying this regularization technique to Example 3.1 results in the solutions presented in Table 3.2.

Table 3.2: Tikhonov regularization example

α	b	x	Ax	$\|Ax - b\|$
1	$(2.000, 2.001)$	$(0.800, 0.800)$	$(1.600, 1.601)$	0.566
1	$(2.000, 2.002)$	$(0.800, 0.800)$	$(1.601, 1.601)$	0.566
0.1	$(2.000, 2.001)$	$(0.997, 0.998)$	$(1.995, 1.996)$	0.007
0.1	$(2.000, 2.002)$	$(0.997, 0.998)$	$(1.996, 1.997)$	0.007
0.001	$(2.000, 2.001)$	$(1.000, 1.000)$	$(2.000, 2.001)$	7×10^{-7}
0.001	$(2.000, 2.002)$	$(0.800, 1.200)$	$(2.000, 2.002)$	5×10^{-4}

If it is assumed the input and output have equal error variances ($\alpha = 1$), then the optimal Tikhonov solution hardly satisfies either of the two original equations. If α is chosen to be too small, then the Tikhonov solution hardly regularizes the problem. One must take care in selecting a weighting factor to balance stability and adherence to the original system of equations.

Instead of adding a new term to regularize the problem, the original system can be reduced to an 'effective', well-conditioned, system of equations, and quadratic program techniques can be employed. In the ill-conditioned example, the two equations are equivalent from a practical point of view, so one of them can be safely removed. All that is required is a method to detect unnecessary equations. This can be accomplished through a *singular value decomposition* (SVD).

3.3.3.2 Singular Value Decomposition

Any matrix of size $m \times n$ can be decomposed into the following form:

$$A = U\Sigma V^{\mathrm{H}},$$

where U and V are unitary matrices, and Σ is a diagonal matrix containing only non-negative real values. This is the singular value decomposition of the matrix A. The diagonal entries of Σ are known as the *singular values* of A, and are typically arranged in a non-increasing order. They are the eigenvalues of the matrix $A^{\mathrm{H}}A$. The rank of A can be defined as the number of non-zero singular values. If a matrix is singular, then at least one of its singular values is zero. An ill-conditioned, non-singular matrix has at least one singular value that is relatively close to zero. This property can be used to define an *effective rank*, which is the number of singular values that are above some threshold. Those values below the threshold can be considered negligible, or effectively zero.

There are properties of the SVD that can help reduce the ill-conditioned system of equations to an effective system. In the field of complex numbers, the columns of U form an orthonormal basis for $\mathbb{C}^m$, and the columns of V form an orthonormal basis for $\mathbb{C}^n$. These basis vectors are arranged in a particular order: the first column of U points the direction in which the columns of A vary the most; the second column of U points in the second most varying direction that is orthogonal to the first; the third points in the third most varying direction that is orthogonal to the first two, and so on. The standard deviations along these directions are given by the corresponding singular values. If the variance along a particular direction is less than, or close to, the noise variance, then this information can be attributed to noise. The noise variance, therefore, sets a threshold below which information can be ignored. If the singular values below the threshold are set to zero, a new matrix is formed:

$$\hat{A} = U\hat{\Sigma}V^{\mathrm{H}},$$

where $\hat{\Sigma}$ contains the truncated set of singular values. It can be shown that if $\mathrm{rank}\{\hat{A}\} = r$, then $\hat{A}$ is the unique matrix of rank r that best approximates A in the Frobenius sense.

Up to this point, only the operator, A, was considered. The new problem is to solve $\hat{A}x = b$, where $\hat{A}$ now has a potentially reduced rank, r. Since the right-hand side was not modified, this could potentially create a set of inconsistent equations. To handle this, consider the original system again:

$$Ax = b \quad \implies \quad \Sigma V^{\mathrm{H}}x = U^{\mathrm{H}}b.$$

The last $m - r$ rows of Σ will be set to zero in $\hat{\Sigma}$, which causes the left-hand sides of these equations to be identically zero. This means the last $m-r$ equations will be inconsistent unless the same rows of $U^{\mathrm{H}}b$ are zero. Therefore, to create a consistent system of r equations, the information in the last $m - r$ rows of this modified system should be discarded. The equivalent system in terms of $\hat{A}$ can be computed:

$$\hat{\Sigma}V^{\mathrm{H}}x = \hat{I}U^{\mathrm{H}}b$$
$$\hat{A}x = \underbrace{U\hat{I}U^{\mathrm{H}}b}_{\hat{b}}$$

where $\hat{I}$ is an $m \times m$ matrix with ones along the first r positions on the diagonal and zeroes everywhere else. In this way, a consistent set of m equations is generated with $n - r$ degrees of freedom. If only r equations are desired, then the system must be left in the form $\hat{\Sigma}V^{\mathrm{H}}x = \hat{I}U^{\mathrm{H}}b$. This method can be applied to Example 3.1 to reduce the set of equations to an effective system.

Example 3.2. *Determine the unknown variables, x_1 and x_2, in the following system of linear equations, by first reducing the problem to an effective system:*

$$1.000\,x_1 + 1.000\,x_2 = 2.000, \tag{3.26}$$
$$1.000\,x_1 + 1.001\,x_2 = 2.001. \tag{3.27}$$

If the reduced system is underdetermined, find the solution with the minimum ℓ_2-norm.

When converted into the form $Ax = b$, the matrix, A, has the following singular value decomposition:

$$U = \begin{bmatrix} -0.7069 & -0.7073 \\ -0.7073 & 0.7069 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2.0005 & 0 \\ 0 & 0.0005 \end{bmatrix} \qquad V = \begin{bmatrix} -0.7069 & -0.7073 \\ -0.7073 & 0.7069 \end{bmatrix}$$

It is quite clear that the second singular value is negligible, so the operator has an effective rank of one. The new, reduced system is

$$\hat{\Sigma} V^{\text{H}} x = \hat{I} U^{\text{H}} b$$

$$\begin{bmatrix} -1.4142 & -1.4149 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = -2.8291 \qquad \Longleftrightarrow \qquad \begin{bmatrix} 1.0000 & 1.0005 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2.0005$$

Therefore, the two equations can be effectively replaced by a single equation:

$$1.0000\, x_1 + 1.0005\, x_2 = 2.0005.$$

Note that this is simply the average of the two original equations. The solution with minimum norm is found using EQP to be

$$(x_1, x_2) = (0.9997, 1.0002).$$

This can be compared to the results from Tikhonov regularization.

In the example, the SVD reduction method was able to identify that the two equations were practically the same, and replaced them the average of the two. Assuming the coefficients are independent and normally distributed with equal variances, this procedure will determine the r equations that best fit the original problem. The method can be generalized further by adding a weighting matrix,

$$WAx = Wb.$$

Just as for generalized least-squares, W can decorrelate noise and equalize noise variances if the statistics are known. Once the system is reduced to a well-conditioned one, a regularized solution can be found with the use of an EQP algorithm.

The major advantage of the SVD method over the Tikhonov one is that it will never distort well-conditioned problems. Even if the system is ill-conditioned, it will always choose solutions that best fit the reduced problem. If the system is overdetermined, the SVD will provide the ordinary least-squares solution. If the system is underdetermined, equality-constrained quadratic programming can find the best solution fitting certain regularization criteria. If the system is inconsistent, the SVD method will find the best approximating consistent system. Optimal regularization is related to the uncertainty in the coefficients, rather than the expected noise in the solution. The columns of V corresponding to zero singular values of A span the nullspace of A, so null-space EQP methods are very easy to implement after decomposition. Finally, the SVD method averages similar equations together, which is probably the most intuitive course of action. For more information on the singular value decomposition, see Trefethen [67] and Golub [19]. Tikhonov and other forms of regularization for inverse problems are described in Tarantola [65].

3.4   Pre-Processing

In any experiment involving measurements, the measured values will include some form of undesired information. This is usually in the form of errors caused by small random fluctuations, called *noise*, but can also be substantial information that is either not applicable to, or captured by, the model. Filtering is a process of removing this unwanted information, leaving a useful dataset for the application.

3.4.1   Noise

All measurements are subject to measurement errors. Physics is subject to small fluctuations. Computers have finite precision. Models are 'big pictures', full of approximations. For these reasons, a measured quantity can never be exact; it is always corrupted by some amount of noise. This is usually modelled by adding a random signal,

$$\hat{y}(t) = y(t) + \eta(t),$$

where $\hat{y}$ is the measured signal, y is the 'true' signal, and η is a random function representing additive noise. As long as all the contributors to this noise are independent and have finite means and variances, the central limit theorem implies that this random function will be approximately normally distributed. If the noise at any two differing times is uncorrelated, then it is said to be *white*. Combining these two qualities results in the standard noise model: *Additive White Gaussian Noise* (AWGN).

Noise is a particular annoyance when derivatives of the data are required. Differentiation can amplify noise, sometimes to the point where the underlying signal is not recognizable. An example of this is shown in Figure 3.12. White noise contains an equal amount of energy at all frequencies, resulting in a flat power spectral density. By taking the derivative of a noisy signal, all components are amplified by a factor equal to their frequency. The higher frequencies in the noise may be amplified high enough to overpower the signal. For this reason, it is important to remove, or filter, the undesired high-frequencies. The most straightforward method of accomplishing this is through a frequency-domain filter.

3.4.1.1   Low-Pass Filter

A low-pass filter allows frequencies below a certain cut-off to pass through, but removes all higher frequencies. This is easily performed in the frequency domain: transform a signal into the frequency domain, set all unwanted high frequencies to zero, then transform the signal back to the time (or space) domain. Care must be taken to not remove important information. If it is known beforehand the the signal has a maximum bandwidth, then all frequencies outside this frequency range can be safely removed.

The design of a low-pass filter is rather simple. The ideal filter has the following frequency response:

$$H(f, f_{\text{low}}) = \begin{cases} 1, & f < f_{\text{low}} \\ 0, & \text{otherwise} \end{cases}$$

All frequencies below the cut-off are kept, while all those above are removed. When implemented for a finite signal, it must be assumed that the data is either periodic, or extended by zeroes in all directions.

(a) Time-domain

(b) Frequency-domain (Periodogram)
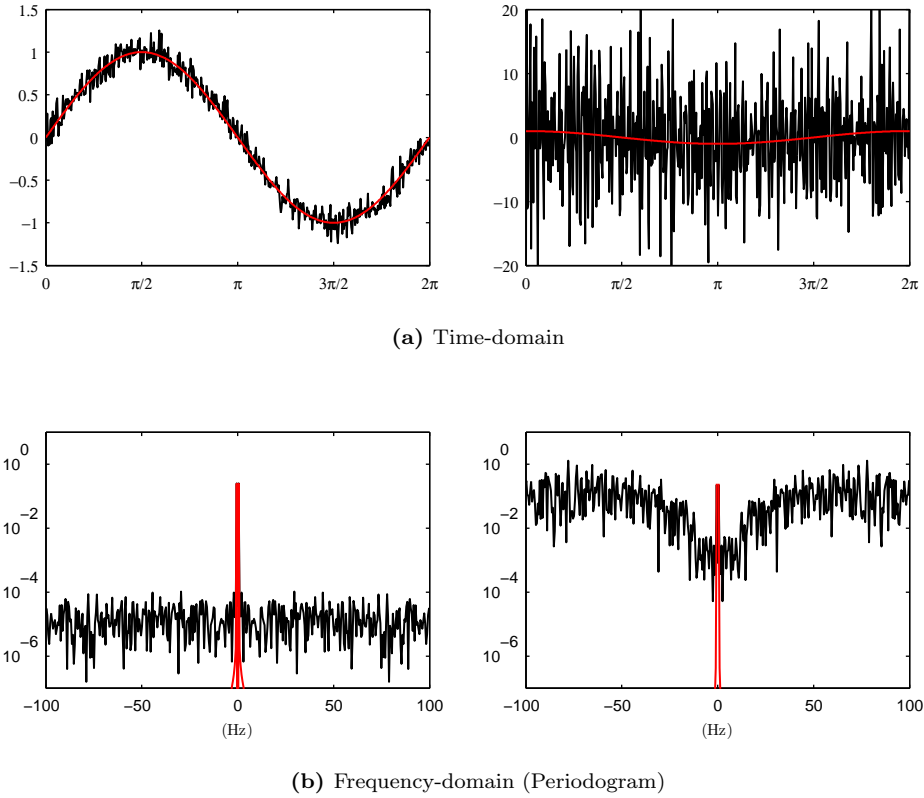
Figure 3.12: Amplification of noise by differentiation. The left image shows the noisy (black) and noise-free (red) signals before differentiation. The underlying trend of the data is clearly captured in both the time and frequency domains. The right image shows the signals after differentiation. The derivative operator amplifies the higher frequency content in the noise, drowning out the signal in the time-domain.

The ideal filter can sometimes introduce oscillations due to the discontinuous nature of the frequency response. To compensate, smooth transitions can be introduced. An example of this is the Butterworth low-pass filter, that has a frequency response given by

$$H(f, f_{\text{low}}) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_{\text{low}}}\right)^{2k}}},$$

where f is the frequency variable, and k is the order of the filter. The higher the order of the filter, the sharper the frequency cut-off is.

Other frequency-domain filters can be constructed from this low-pass filter. A high-pass filter admits frequencies above a certain threshold. A band-pass filter admits frequencies within a certain frequency

range. A stop-band filter admits frequencies outside a certain range. These can be constructed as follows:

$$H_{\text{high}}(f, f_h) = 1 - H_{\text{low}}(f, f_h)$$
$$H_{\text{band}}(f, f_l, f_h) = H_{\text{low}}(f, f_l) \, H_{\text{high}}(f, f_h)$$
$$H_{\text{stop}}(f, f_l, f_h) = 1 - H_{\text{band}}(f, f_l, f_h),$$

where f_l and f_h are the low and high cut-off frequencies, respectively. With these filters, certain unwanted frequencies can be removed from the data.

3.4.1.2 Median Filter

A disadvantage of the low-pass filter is that it smooths out sharp edges. The Fourier transform of a signal containing a discontinuity has energy at all frequencies. To prevent this sharp edge from being smoothed, a filter must be able to remove noise without drastically modifying the frequency content.

A median filter of length N simply replaces every value of a signal with the median of the surrounding N values:

$$\hat{f}(\boldsymbol{x}) = \text{median} \left\{ f(\boldsymbol{x}_i) \mid i = 1 \dots N \right\}$$

This type of filter is very useful in removing outliers. Unlike the low-pass filter, this is applied in the time or space domain. Since it is not based on frequencies, it is able to preserve sharp edges. At the same time, it is able to remove speckle noise.

3.4.1.3 Savitsky-Golay Filter

Another time/space domain set of filters are the Savitsky-Golay (SG) smoothing filters. Oliphant *et al.* use these to remove noise before performing the AIDE elasticity inversion [50]. In the filtering process, it is assumed that the function can be locally approximated by a low-order polynomial of the form

$$f(\boldsymbol{x}) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} a_{i_1 i_2 \cdots i_n} (x_1)^{i_1} (x_2)^{i_2} \dots (x_n)^{i_n}, \qquad \sum_{k=1}^{n} i_k \leq O$$

where O is the maximum order, n is the dimension of the space, and a_j are coefficients to be determined. This can be expressed in vector form:

$$f(\boldsymbol{x}) = \boldsymbol{P}(\boldsymbol{x})^{\text{H}} \boldsymbol{a},$$

where $\boldsymbol{P}$ is the vector of polynomials $\left[(x_1)^{i_1} \dots (x_n)^{i_n} \right]$, and $\boldsymbol{a}$ is the vector of coefficients. If this function is evaluated at N points, where N is larger than the degrees of freedom in f, then a set of linear equations are created:

$$\begin{bmatrix} f(\boldsymbol{x}_1) \\ f(\boldsymbol{x}_2) \\ \vdots \\ f(\boldsymbol{x}_N) \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}(\boldsymbol{x}_1)^{\text{H}} \\ \boldsymbol{P}(\boldsymbol{x}_2)^{\text{H}} \\ \vdots \\ \boldsymbol{P}(\boldsymbol{x}_N)^{\text{H}} \end{bmatrix} \boldsymbol{a},$$

which has a weighted least-squares solution,

$$\hat{a} = (PWP^{\text{H}})^{-1} PW\boldsymbol{f}.$$

The vector $\boldsymbol{f}$ consists of the measured noisy data, and W is a matrix of weights that can be used to decorrelate noise or favour central points. If the N points surround some central point, $\boldsymbol{x}_o$, then the determined polynomial can be thought to represent this point. The smoothed version of the function at $\boldsymbol{x}_o$ can be determined using the best-fitting coefficients:

$$\hat{f}(\boldsymbol{x}_o) = \boldsymbol{P}(\boldsymbol{x}_o)^{\text{H}}\hat{a}$$
$$= \boldsymbol{P}(\boldsymbol{x}_o)^{\text{H}} (PWP^{\text{H}})^{-1} PW\boldsymbol{f}.$$

The matrix of weights is usually a diagonal matrix designed to favour points closest to the central point. Note that this equation simply expresses the smoothed version of f as a linear combination of the measured values surrounding $\boldsymbol{x}_o$:

$$\hat{f}(\boldsymbol{x}_o) = \sum_{k=1}^{N} b_k f(\boldsymbol{x}_k),$$

which is the form of a *Finite Impulse Response* (FIR) filter. Thus, as long as the grid spacing is uniform, the coefficients need only be computed at $\boldsymbol{x}_o = 0$, and the smoothed version, $\hat{f}$, can be computed via a convolution.

This technique also lends itself well to approximating derivatives. Since the coefficients determined at each point describe the best-fitting polynomial of degree O, the derivative at this point can be approximated by the derivative of the polynomial:

$$\frac{\partial^j f}{\partial x_1^{j_1} \partial x_2^{j_2} \cdots \partial x_n^{j_n}} = \boldsymbol{Q}^{\text{H}}\hat{a}, \qquad j = j_1 + \ldots + j_n,$$

where $\boldsymbol{Q}$ is the vector composed of the appropriate derivatives of $\boldsymbol{P}$:

$$\boldsymbol{Q}(\boldsymbol{x}) = \left[\left(\frac{\partial^{j_1}}{\partial x_1^{j_1}} (x_1)^{i_1} \right) \cdots \left(\frac{\partial^{j_n}}{\partial x_n^{j_n}} (x_n)^{i_n} \right) \right]_{i_1 \ldots i_n}.$$

Again, this expresses the derivative of the approximating polynomial at $\boldsymbol{x_o}$ as a linear combination of the values at the surrounding points:

$$\frac{\partial^j f(\boldsymbol{x}_o)}{\partial x_1^{j_1} \partial x_2^{j_2} \cdots \partial x_n^{j_n}} = \boldsymbol{Q}(\boldsymbol{x}_o)^{\text{H}} (PWP^{\text{H}})^{-1} PW\boldsymbol{f}$$
$$= \sum_{k=1}^{N} c_k f(\boldsymbol{x}_k),$$

which can be performed via a convolution. In this manner, a set of n-dimensional filters can be constructed to not only smooth a function, but also estimate its smoothed derivatives.

Example 4. *Determine the second-order, one-dimensional Savitsky-Golay smoothing filter and first two derivative filters, using a 5-point stencil.*

A second-order polynomial has the form

$$f(x) = a_1 + a_2 x + a_3 x^2.$$

The polynomial matrix and evaluation vectors are constructed as follows:

$$P^{\text{H}} = \begin{bmatrix} 1 & -2\Delta x & (-2\Delta x)^2 \\ 1 & -1\Delta x & (-1\Delta x)^2 \\ 1 & 0 & 0 \\ 1 & 1\Delta x & (1\Delta x)^2 \\ 1 & 2\Delta x & (2\Delta x)^2 \end{bmatrix} \qquad \begin{aligned} \boldsymbol{P}(0)^{\text{H}} &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \\ \boldsymbol{Q}_1(0)^{\text{H}} &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\ \boldsymbol{Q}_2(0)^{\text{H}} &= \begin{bmatrix} 0 & 0 & 2 \end{bmatrix} \end{aligned}$$

where $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are used to find the first and second derivatives at the central point, respectively. The corresponding SG filters are:

$$\hat{f}(x) = \left[-\frac{3}{35}f(x - 2\Delta x) + \frac{12}{35}f(x - \Delta x) + \frac{17}{35}f(x) + \frac{12}{35}f(x + \Delta x) - \frac{3}{35}f(x + 2\Delta x) \right]$$

$$\frac{\partial}{\partial x}\hat{f}(x) = \frac{1}{\Delta x}\left[-\frac{1}{5}f(x - 2\Delta x) - \frac{1}{10}f(x - \Delta x) + 0f(x) + \frac{1}{10}f(x + \Delta x) + \frac{1}{5}f(x + 2\Delta x) \right]$$

$$\frac{\partial^2}{\partial x^2}\hat{f}(x) = \frac{2}{\Delta x^2}\left[\frac{1}{7}f(x - 2\Delta x) - \frac{1}{14}f(x - \Delta x) - \frac{1}{7}f(x) - \frac{1}{14}f(x + \Delta x) + \frac{1}{7}f(x + 2\Delta x) \right].$$

These filters are expressed in the time (or space) domain, but can easily be implemented in the frequency domain by finding the appropriate frequency response.

3.4.1.4   Relaxation

An alternative to using explicit filters to remove noise is to allow the data to evolve, or relax, into a smoother state. This can be accomplished by finding a new function that balances between matching the data and fitting some regularity conditions. This is the approach taken by Park *et al.* [54]. In the finite-element method, the relaxed data, $\hat{\boldsymbol{u}}$, is determined by minimizing

$$F(\hat{\boldsymbol{u}}) = \alpha_1 \int_\Omega \|\hat{\boldsymbol{u}} - \boldsymbol{u}\|^2 dV + \alpha_2 \int_\Omega (\nabla \cdot \hat{\boldsymbol{u}})^2 dV + \alpha_3 \int_\Omega \|\nabla \hat{\boldsymbol{u}}\|^2 dV,$$

where Ω is the entire volume, $\boldsymbol{u}$ is the measured data, and $\{\alpha_i\}$ control the relative weightings between the factors in the minimization. These weights were chosen such that $\alpha_2 \gg \alpha_1 \gg \alpha_3$, so that incompressibility was the dominating regularization factor. This reflects the knowledge that soft tissues are relatively incompressible.

With this technique, the smoothed version of the displacements can be regularized to satisfy certain conditions, such as incompressibility. This is not possible with the previously mentioned filters. Care must be taken in choosing the weighting factors, however, or else the smoothed version may over-fit the regularity conditions, and not fit the original data. This is the same issue that is present in Tikhonov regularization.

In order to implement this minimization on a grid of measurements, the operators can be discretized and represented by matrices:

$$F(\hat{\boldsymbol{u}}) = \alpha_1 \left\| D_{\text{div}}\hat{\boldsymbol{u}} \right\|^2 + \alpha_2 \left\| D_{\text{grad}}\hat{\boldsymbol{u}} \right\|^2 + \alpha_3 \left\| \hat{\boldsymbol{u}} - \boldsymbol{u} \right\|^2.$$

Here, $\boldsymbol{u} \in \mathbb{C}^n$ is a rasterized vector of the harmonic displacement field. This rasterization is performed as follows:

$$\boldsymbol{u}\Big(j + (k-1)N_x + (l-1)N_xN_y + (i-1)N_xN_yN_z\Big) = u_i(j,k,l),$$

where u_i is the ith component of the three-dimensional measured field, defined on a grid with dimensions $N_x \times N_y \times N_z$. The minimization can be generalized by adding weighting matrices. In fact, by adding weights, this technique can be seen in a new light. Minimization over the entire domain may require the inversion of a very large matrix. A weighting matrix can be used to select a smaller, manageable window of the data. The weighted least-squares solution within this window is then given by

$$\hat{\boldsymbol{u}} = \alpha_1 \left(\alpha_1 W + \alpha_2 D_{\text{div}}^{\text{H}} W D_{\text{div}} + \alpha_3 D_{\text{grad}}^{\text{H}} W D_{\text{grad}}\right)^{-1} W \boldsymbol{u}.$$

Assume W is only non-zero along diagonal entries corresponding to points about some central point, $\boldsymbol{x}_o$. If there are N points chosen around $\boldsymbol{x}_o$, then W has nN non-zero elements (N for each component of the displacement field). Then, similar to the Savitsky-Golay derivation, the smoothed version, $\hat{\boldsymbol{u}}(\boldsymbol{x}_o)$, can be expressed as a linear combination of the measured field at these surrounding points:

$$\hat{u}_i(\boldsymbol{x}_o) = \sum_{j=1}^{n} \sum_{k=1}^{N} c_{jk} u_j(\boldsymbol{x}_k).$$

These entries can be determined by multiplying $\hat{\boldsymbol{u}}$ by the appropriate $n \times N^n$ matrix, I_o, that singles out the entries corresponding to $u_i(\boldsymbol{x}_o)$. The smoothed field at the central point is given by:

$$\hat{\boldsymbol{u}}(\boldsymbol{x}_o) = \left[\alpha_1 I_o \left(\alpha_1 W + \alpha_2 D_{\text{div}}^{\text{H}} W D_{\text{div}} + \alpha_3 D_{\text{grad}}^{\text{H}} W D_{\text{grad}}\right)^{-1} W\right] \boldsymbol{u}.$$

In this way, a filter is derived to perform the relaxation process,

$$h = \left[\alpha_1 I_o \left(\alpha_1 W + \alpha_2 D_{\text{div}}^{\text{H}} W D_{\text{div}} + \alpha_3 D_{\text{grad}}^{\text{H}} W D_{\text{grad}}\right)^{-1} W\right].$$

This filter can be re-shaped, creating n^2 separate filters, such that:

$$\hat{u}_i = \sum_{j=1}^{n} h_{ij} * u_j,$$

where each h_{ij} contains N values, and has the same grid dimensions as the windowed region. The difference between this filter set and the previous ones is that there are coupling terms between all the components of the vector field. The other filters act on each component of the field separately. This may help eliminate noise it is uncorrelated between components of the measured field. Two-dimensional relaxation filters are constructed in the following example.

Example 5. *Consider a vector field $\boldsymbol{u} : \mathbb{R}^2 \to \mathbb{C}^2$. Construct a pair of relaxation filters using a 3×3 stencil, with grid spacings $\Delta x = \Delta y = 1$, and parameters $\alpha_1 = \alpha_2 = \alpha_3 = 1$.*

The field must first be rasterized. Arrange the points such that $\boldsymbol{x} = [(x_1, y_1)\ (x_2, y_1)\ \ldots\ (x_3, y_3)]^{\mathrm{T}}$, and $\boldsymbol{u} = [u_1^{\mathrm{H}}\ u_2^{\mathrm{H}}\ u_3^{\mathrm{H}}]^{\mathrm{H}}$. The operator matrices are the following:

$$D_{\mathrm{grad}} = \begin{bmatrix} D_x & 0 \\ 0 & D_y \end{bmatrix} \qquad D_{\mathrm{div}} = \begin{bmatrix} D_x & D_y \\ 0 & 0 \end{bmatrix} \qquad W = I,$$

where D_x and D_y are block matrices that approximate the derivative of a discrete function. In this case, second-order finite difference matrices are constructed, with upwinding and downwinding on the boundaries. The matrix that isolates the central point is given by

$$I_o = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Performing the inversion and reshaping the filters into 3×3 matrices yields:

$$h_{11} = \begin{bmatrix} -0.0010 & 0.3190 & -0.0010 \\ 0.0020 & 0.3620 & 0.0020 \\ -0.0010 & 0.3190 & -0.0010 \end{bmatrix}, \qquad h_{12} = \begin{bmatrix} 0.0122 & 0 & -0.0122 \\ 0 & 0 & 0 \\ -0.0122 & 0 & 0.0122 \end{bmatrix},$$

$$h_{21} = \begin{bmatrix} 0.0122 & 0 & -0.0122 \\ 0 & 0 & 0 \\ -0.0122 & 0 & 0.0122 \end{bmatrix}, \qquad h_{22} = \begin{bmatrix} -0.0010 & 0.3190 & -0.0010 \\ 0.0020 & 0.3620 & 0.0020 \\ -0.0010 & 0.3190 & -0.0010 \end{bmatrix}.$$

The resulting smoothed, or relaxed, displacement field is given by

$$\hat{u}_1 = h_{11} * u_1 + h_{12} * u_2$$
$$\hat{u}_2 = h_{21} * u_1 + h_{22} * u_2.$$

Note that the result is nothing but a weighted average of all the surrounding points. The gradient and divergence terms couple the equations, so that $\hat{u}_i$ depends on all $\{u_j\}$. If $\alpha_2 = \alpha_3 = 0$, then the relaxation filters degenerate to $h_{ij} = \delta_{ij}$, and all smoothing function is lost.

The relaxation filter developed in this section was designed to regularize the data by minimizing the gradient and the divergence. The same derivation can be repeated for any operator, as long as it is a local operator and can be expressed in matrix form.

A more detailed description of noise filtering techniques can be found in Proakis [56], and for the specific application of image processing, in Gonzalez and Woods [20].

All the noise filtering techniques mentioned in this section are compared in Figure 3.13. A discontinuity is introduced in the noise-free signal to test how this is treated by the various filtering methods. Filtering is important for direct inversion methods in MRE, since derivatives of the noisy data are required, and differentiation amplifies noise. In the figure, it is apparent that the median filter best captures the discontinuity. This was expected; median filters are designed to filter out noise without affecting edges. The low-pass filter results in the smoothest estimate and derivative, however, these are

(a) Unfiltered

(b) Median

(c) Low-pass, $f_{\text{low}} = 10$Hz

(d) Savitsky-Golay, 2nd Order

(e) Relaxation, 1st Derivative
$\alpha = (1, 0.1, 0)$

(f) Relaxation, 2nd Derivative
$\alpha = (1, 0.1, 0)$
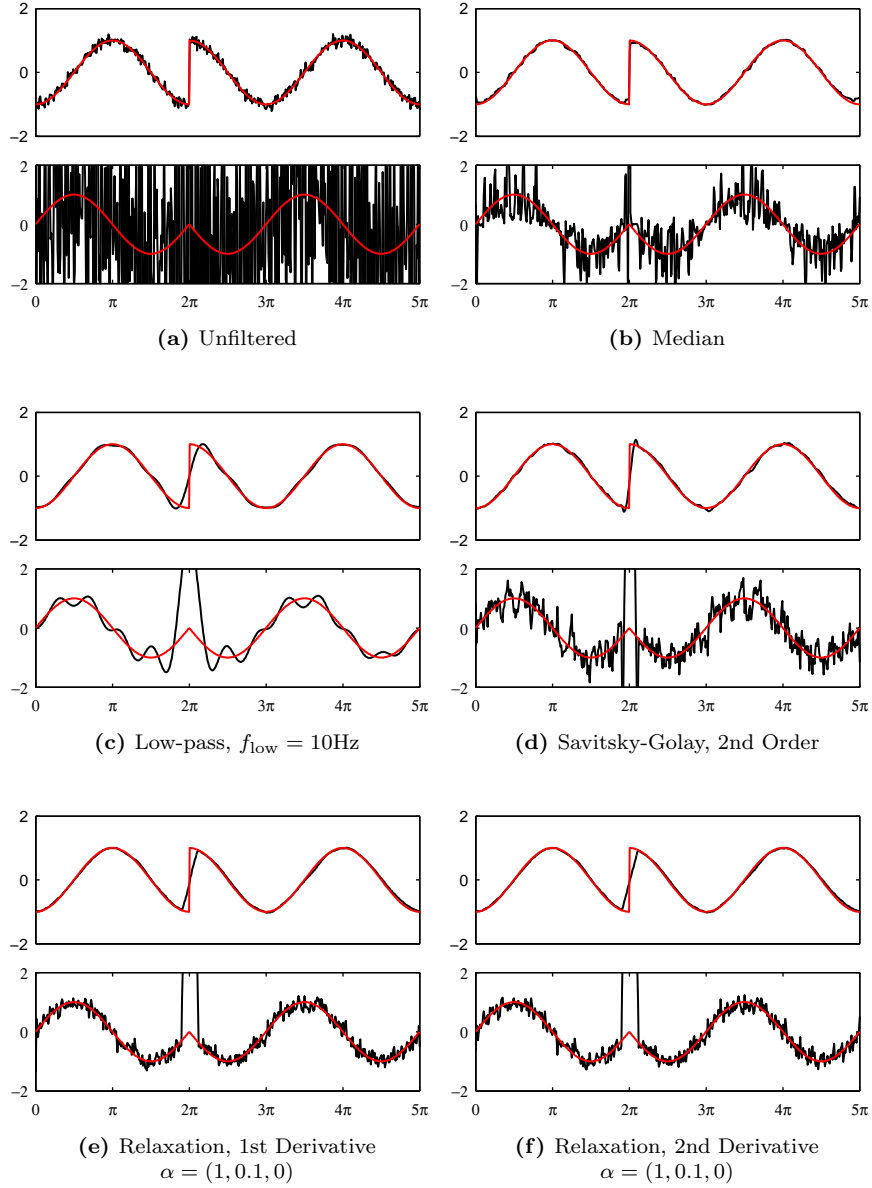
Figure 3.13: Filtering a noisy signal. The unfiltered signal is designed to have a jump discontinuity and an SNR of 20 dB. The signal is then smoothed with the labelled smoothing filters (top images). The grid size is 512 points, and the window size of the median, SG and relaxation filters is 21 points. The derivative of the smoothed data is estimated using fourth-order centered finite differencing (bottom images). Ideal solutions are shown in red.

also the most inaccurate. Discontinuities cannot be captured without high-frequency information. The Savitsky-Golay filter perform relatively well. It is only outperformed by the median filter in capturing the jump discontinuity. The relaxation filters seem to produce the smoothest results. Again, this is not surprising, since the regularization condition forces the norm of the derivatives to be minimized. Unfortunately, for the relaxation filters to work well, a rather large data window is required due to the discrete derivative approximations. This made them unsuitable for three-dimensional MRE.

3.4.1.5   Matched Filter

The median, low-pass, Savitsky-Golay and relaxation filters all attempt to clean a signal by removing the unwanted noise. Matching filters work on a different principle. Instead of attempting to remove noise, one can try to minimize the effect of the noise on the output.

Assume a discrete signal is corrupted by additive noise:

$$\hat{u} = u + \sigma W \eta,$$

where $\hat{u}$ is the measured signal in vector form, u is the noise-free signal, η is white Gaussian noise with unit variance, σ is a factor to scale the variances, and W is an invertible matrix that models correlation and unequal variances in the noise. If the noise is uncorrelated its components have equal variances, then $W = I$. Otherwise, the noise has a correlation matrix given by

$$\begin{aligned} \Sigma_\eta &= E\left\{ \sigma W \eta \left(\sigma W \eta\right)^{\text{H}} \right\} \\ &= \sigma^2 W \; E\left\{\eta\eta^{\text{H}}\right\} \; W^{\text{H}} \\ &= \sigma^2 W W^{\text{H}}, \end{aligned}$$

where E is used to denote the expected value. When the measured signal is acted on by a correlation filter, h, the output is given by

$$\begin{aligned} \hat{y} &= h^{\text{H}} \hat{u} \\ &= h^{\text{H}} u + h^{\text{H}} \sigma W \eta \\ &= y_u + y_\eta. \end{aligned}$$

The output signal-to-noise ratio (SNR) is the following:

$$\text{SNR} = \frac{|y_u|^2}{E\left\{|y_\eta|^2\right\}}.$$

In order to minimize the effect of noise on the output function, the filter, h, should maximize this SNR.

This can easily be determined by expanding the expression:

$$
\begin{aligned}
\text{SNR} &= \frac{y_u y_u^{\text{H}}}{E\left\{y_\eta y_\eta^{\text{H}}\right\}} \\
&= \frac{h^{\text{H}} u u^{\text{H}} h}{\sigma^2 h^{\text{H}} W W^{\text{H}} h} \\
&= \frac{h^{\text{H}} W W^{-1} u u^{\text{H}} \left(W^{\text{H}}\right)^{-1} W^{\text{H}} h}{\sigma^2 h^{\text{H}} W W^{\text{H}} h}.
\end{aligned}
$$

The Cauchy-Schwarz inequality states $(a^{\text{H}} b) b^{\text{H}} a \leq (a^{\text{H}} a)(b^{\text{H}} b)$, with equality only if $a = \alpha b$ for some $\alpha \in \mathbb{R}$. Applying this to the previous expression for SNR,

$$
\begin{aligned}
\text{SNR} &\leq \frac{(h^{\text{H}} W W^{\text{H}} h) \left(u^{\text{H}} [W^{\text{H}}]^{-1} W^{-1} u\right)}{\sigma^2 h^{\text{H}} W W^{\text{H}} h} \\
&= \frac{u^{\text{H}} \Sigma^{-1} u}{\sigma^2},
\end{aligned}
$$

with equality only if

$$
W^{\text{H}} h = \alpha W^{-1} u
$$
$$
\implies \quad h = \alpha \Sigma^{-1} u.
$$

For uncorrelated noise, $\Sigma = \Sigma^{-1} = I$. Therefore, in order to minimize the effect of noise on the output, the filter should be matched to the data over the support of the filter. Note that this filter does not help recover a corrupted signal, so cannot be used for smoothing like the previously described filters. For cases where the noise-free signal is known, the matched filter instead maximizes the signal-to-noise ratio in the output of a filtering process. When the noise-free signal is not known, it must be estimated from the noisy version, perhaps by using a smoothing filter.

3.4.2   Directional Filter

In the magnetic resonance elastography experiment, mechanical excitation is typically applied to one side of the tissue, and the majority of the shear waves travel in a direction away from the source. For example, if forcing is applied to the top of a 3D cube of material, most waves will travel from top to bottom. If the waves encounter a boundary or a region of stiffer tissue, however, the waves will refract and reflect. This can create complicated interference patterns. Manduca *et al.* [39] argued that these interference patterns can cause regions with near-zero displacements. In these regions, the signal is dominated by noise, making inversion for the shear modulus unstable. They proposed filtering the waves so that motions in particular directions can be separated. The shear modulus can then be reconstructed by weighting combinations of inversions performed on waves travelling in different directions.

The idea of the directional filter originates with Knutsson *et al.* [29], who used them to estimate local frequencies. The original filters had no time-dependence, however. Manduca *et al.* modified this filter to include the temporal domain.

The form of the filter is best derived in the frequency domain. Consider a spatial direction with unit direction vector $\hat{\boldsymbol{n}}$. All waves oriented with a spatial frequency, $\boldsymbol{k}$, not in the direction of $\hat{\boldsymbol{n}}$, should be

(a) Unfiltered   **(b)** $\cos^2(\theta)$ **(c)** $\cos^{10}(\theta)$
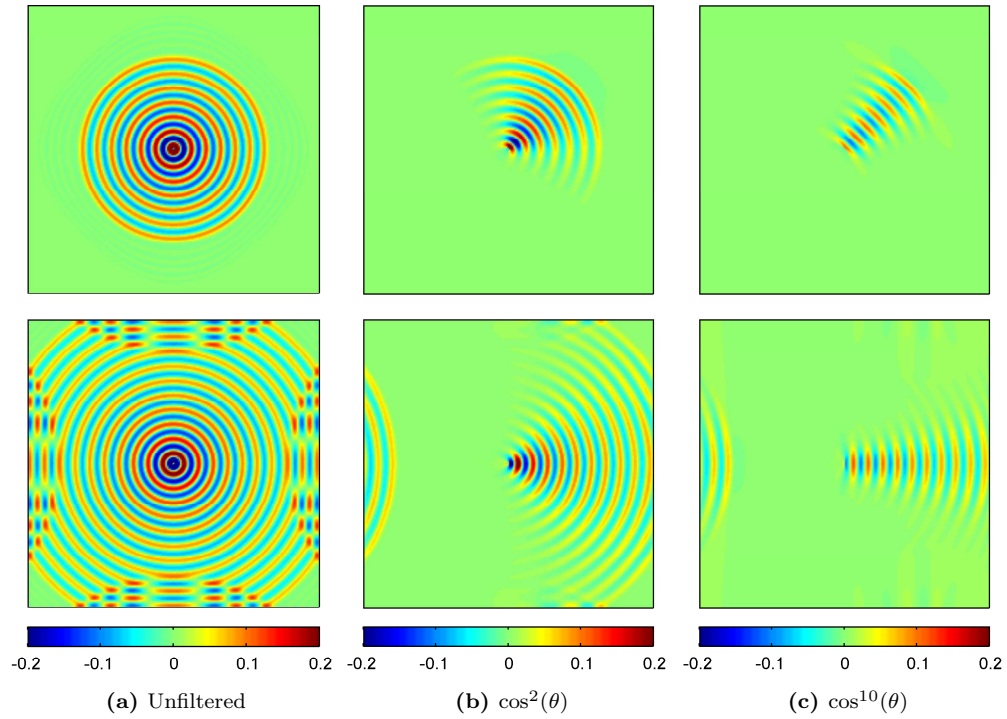
Figure 3.14: Directional filtering of a simulated two-dimensional radial wave, with reflecting boundary conditions. Attenuation is dependent on the angle of propagation relative to the specified direction.

attenuated. A simple function that satisfies this is the vector dot-product. Waves travelling precisely in the direction of $+\hat{\boldsymbol{n}}$ as time increases should remain unaffected, whereas waves travelling in the direction of $-\hat{\boldsymbol{n}}$ as time increases should be completely removed. This leads to the following frequency-domain filter:

$$D_{\hat{\boldsymbol{n}}}^{\alpha}(\boldsymbol{k}, f) = \begin{cases} \left(\dfrac{\boldsymbol{k} \cdot \hat{\boldsymbol{n}}}{\|\boldsymbol{k}\|}\right)^{\alpha}, & \text{if } -(\boldsymbol{k} \cdot \hat{\boldsymbol{n}})f > 0 \\ 1, & \text{if } \|\boldsymbol{k}\| = 0 \\ 0, & \text{otherwise} \end{cases}$$

Any wave with a component propagating in the direction of $-\hat{\boldsymbol{n}}$ follows under the 'otherwise' category, and is completely removed. If a wave makes an angle with the specified direction, $\theta \in [-\pi/2, \pi/2]$, then

the attenuation is given by

$$
\begin{aligned}
\mathrm{Att}(\boldsymbol{k}) &= \left(\frac{\boldsymbol{k} \cdot \hat{\boldsymbol{n}}}{\|\boldsymbol{k}\|} \right)^{\alpha} \\
&= \left(\frac{\|\boldsymbol{k}\| \|\hat{\boldsymbol{n}}\| \cos(\theta)}{\|\boldsymbol{k}\|} \right)^{\alpha} \\
&= \cos^{\alpha}(\theta).
\end{aligned}
$$

Thus, waves travelling entirely in the direction of $\hat{\boldsymbol{n}}$ are unaffected, but other waves are attenuated by a factor of $\cos^{\alpha}(\theta)$. In this way, interference patterns can be simplified by removing all motions but those propagating in a single given direction.

An example of directional filtering is presented in Figure 3.14. For this example, a two-dimensional radial wave is generated on a square domain, with reflecting (zero-Neumann) boundary conditions. Two snapshots in time are taken: one before the wave encounters the boundaries, and one after. Two different filtering directions were chosen, with different values of α. Notice that the interference patterns caused by reflections are completely removed. The parameter, α, can be used to tune the range of directions to keep.

In order to apply directional filtering to shear modulus inversion, one must be aware of the consequences of such a filtering technique. The resulting displacement field is no longer physically realistic. Manduca *et al.* acknowledge this, but incorrectly state that since filtering is a linear operation, the filtered data satisfies the same equations of motion.

The filtered displacement field is only guaranteed to satisfy the original system of equations if $\mathbf{U}$ can formally be replaced by the filtered version, $\hat{\mathbf{U}} = h * \mathbf{U}$. However, the presence of spatially varying parameters prevents this. Although convolution is a linear operator, $h * (fu) \neq (h * u)f$, for some spatially varying function, f, in general. Also, $h * (u + f) \neq h * u + f$ for any non-zero f. In order to formally introduce a filter, a convolution with h must be applied to both sides of the original system. The presence of any spatially-varying parameters or forcing terms prevents the filter from solely acting on the data. Therefore, a formal replacement of $\mathbf{U}$ with $\hat{\mathbf{U}}$ is not guaranteed to be valid.

This effect of the replacement of $\mathbf{U}$ with $\hat{\mathbf{U}}$ can be explored for the particular equations involved in MRE. Consider the system with no longitudinal component:

$$
-\omega^2 \mathbf{U} = \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right].
$$

Applying the convolution filter to both sides of the equation maintains validity:

$$
\begin{aligned}
-\omega^2 h * \mathbf{U} &= h * \nabla \cdot \left[\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right] \\
-\omega^2 \hat{\mathbf{U}} &= \nabla \cdot \left[h * \left(\mathcal{M} \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right) \right].
\end{aligned}
\tag{3.28}
$$

On the other hand, if $\mathbf{U}$ is simply replaced by $\hat{\mathbf{U}}$, the following is obtained:

$$
\begin{aligned}
-\omega^2 \hat{\mathbf{U}} &= \nabla \cdot \left[\mathcal{M} \left(\nabla \hat{\mathbf{U}} + \nabla^{\mathrm{T}} \hat{\mathbf{U}} \right) \right] \\
-\omega^2 \hat{\mathbf{U}} &= \nabla \cdot \left[\mathcal{M} \, h * \left(\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right) \right].
\end{aligned}
\tag{3.29}
$$

The two equations, (3.28) and (3.29), are equivalent if and only if

$$\nabla \cdot \iiint_{\Omega} \left[\mathcal{M}(\boldsymbol{x}) - \mathcal{M}(\boldsymbol{y})\right] h(\boldsymbol{x} - \boldsymbol{y})F(\boldsymbol{y}) \, d\boldsymbol{y} = 0,$$

where $F = \nabla \mathbf{U} + \nabla^{\mathsf{T}} \mathbf{U}$, and Ω is the support of the filter. For this to hold for an arbitrary field, F, it can be shown that the following condition must be satisfied:

$$\mathcal{M}(\boldsymbol{x}) - \mathcal{M}(\boldsymbol{y}) = 0, \quad \forall \, \boldsymbol{x}, \boldsymbol{y} \in \Omega.$$

In other words, $\mathcal{M}$ must be constant within the support of the filter. This is again the local homogeneity assumption. The validity of the assumption depends on the variability of $\mathcal{M}$ and the size of the support of h. Therefore, a filtered version of the data is only expected to satisfy the original equations of motion if all parameters can be considered approximately constant over regions as large as the support of the filter.
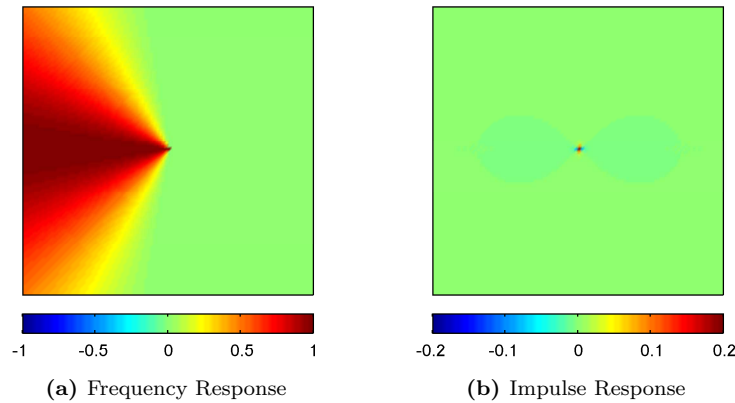

(a) Frequency Response   **(b)** Impulse Response

Figure 3.15: Impulse and frequency responses of the directional filter in two spatial dimensions. The left image shows the frequency response evaluated at $f = 0$. The right image shows the impulse response in the spatial domain evaluated at $t = 0$. Only waves traveling to the right are admitted.

For the directional filter, the impulse response of $D_{\hat{\boldsymbol{n}}}^{\alpha}$ decays quite rapidly, as shown in Figure 3.15. This small support is partially why the original filters are useful in estimating local frequencies. Still, the local homogeneity assumption must be explicitly made in order to argue that the filtered data should satisfy the original equations of motion. This assumption is expected to cause blurring or smearing in regions where the true shear modulus varies sharply. The effects of the directional filter are explored experimentally in Chapter 7.

This necessary condition on the support of the filter does not apply to noise filters. The equations of motion hold for displacement fields without noise. Filtering the noise is not expected to change the nature of the data; it just cleans the data so that the model is more applicable.

3.5   Summary

The estimation of an elasticity parameter in a model from measured displacements is an inverse problem. Such problems are typically ill-posed, which means that a solution does not formally exist, is not unique, or is too sensitive to variations in parameters to be reliable. If a discrete problem satisfies this last condition, it is known as being ill-conditioned.

When no formal solution exists, a 'best-fitting' one can usually be found. This is commonly accomplished through least-squares methods, such as ordinary least-squares (OLS), generalized least-squares (GLS), and feasible generalized least-squares (FGLS). These methods attempt to minimize the squared error in discrete problems, and can be shown to minimize the effect of noise in systems of linear equations.

When no unique solution exists, either outside information is required, or assumptions about the desired solution must be made in order to isolate a particular one. This is known as regularization. One method to accomplish this is to re-pose the original problem in a quadratic minimization framework. Uniqueness can be guaranteed if a convex functional is introduced, of which the desired solution is the minimizing argument. For discrete problems, equality-constrained quadratic programming techniques can be used to isolate particular solutions in large solution spaces.

There are cases when a formal, unique solution does exist, but is too sensitive to small perturbations in the data to be considered reliable. These types of problems also require regularization. This can be accomplished by Tikhonov regularization, in which a stabilizing term is added to the system. The stabilizing term can drastically reduce the sensitivity of the solution, but can also steer it away from satisfying the original problem. An alternative is to reduce the determined system to an effective underdetermined system, and use quadratic optimization methods to regularize. System reduction can be performed by exploiting properties of singular value decompositions. With this method, the final solution is guaranteed to closely satisfy the original problem.

Sensitivity to small variations in the data is largely a problem because of noise. All measurements have sources of error, which is usually modelled as a random component with some assumed distribution. The most common model is additive white Gaussian noise. Noise in the data is particularly an issue in MRE because high-order derivatives of the displacement field appear in the equations of motion. Differentiation amplifies the high-frequency components of noise, which can sometimes drown out the true signal. For this reason, the data is often pre-processed using filtering techniques. The median, low-pass, and Savitsky-Golay filters are designed to remove noise from a signal, while the matched filter is designed to minimize the effect of noise on the output of the filter. Noise can also be removed by allowing the data to relax into a smoother state, according to some regularization conditions. In this way, the effect of the unknown noise signal can be minimized.

The last filter described in this chapter is the directional filter. It is designed for the MRE experiment to remove unwanted displacements caused by reflections and refractions. If the forcing of the tissue is applied to one boundary, most motion will be in a direction away from this boundary. The directional filter attenuates all motions not in the specified direction. This changes the nature of the data, so care must be taken to ensure this does not affect the validity of the shear modulus estimate.

One should be cautious when filtering data, particularly for clinical applications. It may be better to produce a slightly noisy image than to smooth it to the point where important information might be missed. For the application of MRE, some information is known beforehand. Ideally, soft tissues should not contain sharp discontinuities, and certain ranges of stiffness are expected. This limits the range of

expected spatial frequencies in shear motions, which can be used to design appropriate band-pass filters that remove undesired information.

In this chapter, the full tissue model was derived, the form of data and its acquisition were described, methods of pre-processing the data were developed, and frameworks for solving general inverse problems were presented. The stage is now set for the introduction of inversion techniques specific to magnetic resonance elastography.

CHAPTER 4

ALGEBRAIC INVERSION OF THE DIFFERENTIAL EQUATION

Algebraic Inversion of the Differential Equation (AIDE) is a simple, yet accurate direct method for elasticity estimation. Since it is the inspiration for the novel techniques developed in this dissertation, the details are reviewed in this chapter. A series of modifications to the original technique are also discussed, that improve the robustness of the method.

4.1 Original AIDE

The AIDE technique was first published in 1999 by Oliphant *et al.* [50]. Up to that point, all inversion methods were either iterative, or made an assumption that removed the coupling between the three polarizations of motion. The method began with the original equations of motion for a linearly viscoelastic isotropic medium,

$$-\rho_o\,\omega^2\mathbf{U} = \nabla\!\cdot\!\left[\mathcal{M}\left(\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right)\right] + \nabla\left[\Lambda\,\nabla\!\cdot\!\mathbf{U}\right].$$

If it is assumed that the displacements field, $\mathbf{U}$, is given, then this is a set of three differential equations for $\mathcal{M}$ and Λ. Without knowing boundary conditions, the system is difficult to solve; regularization conditions may be required. To avoid this issue, the local homogeneity assumption was proposed: on the order of the scan resolution, the unknown parameters can be considered locally constant. As described in Section 3.1.8, this leads to the following simplified equations of motion:

$$-\rho_o\,\omega^2\mathbf{U} = \mathcal{M}\,\nabla^2\mathbf{U} + (\Lambda + \mathcal{M})\,\nabla\nabla\!\cdot\!\mathbf{U}. \tag{4.1}$$

Equation (4.1) is an algebraic equation for ρ_o, $\mathcal{M}$ and Λ, which can be solved using algebraic techniques. The name of the method is slightly misleading, since the differential equation is not truly inverted algebraically; rather, the differential equation is approximated by an algebraic one, and this is solved. A true algebraic method for inverting the differential equation is presented in Chapter 6.

Since the displacement field is three-dimensional, the system in Equation (4.1) is a set of three equations. Therefore, as long as the system is consistent, determined, and well-conditioned, all three unknown parameters can be identified. In order to improve stability, however, the problem was reduced to two variables. Either ρ_o was considered to be equal to the density of water, $1\ \mathrm{g/cm^3}$, or else the ratios $\mathcal{M}/\rho_o$ and Λ/ρ_o were estimated. The reduced system can be written in matrix form,

$$\underbrace{\begin{bmatrix} U_{1,ii} & U_{i,i1} \\ U_{2,ii} & U_{i,i2} \\ U_{3,ii} & U_{i,i3} \end{bmatrix}}_{=A} \begin{bmatrix} \mathcal{M} \\ \Lambda + \mathcal{M} \end{bmatrix} = -\omega^2 \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix},$$

73

which has a least-squares solution

$$\begin{bmatrix} \mathcal{M} \\ \Lambda + \mathcal{M} \end{bmatrix} = -\omega^2 \left(A^{\mathrm{H}} A \right)^{-1} A^{\mathrm{H}} \mathbf{U}.$$

All that is required to estimate the shear modulus is the matrix of derivative terms, A. To prevent the amplification of noise by differentiation, the data was smoothed using Savitsky-Golay filters. These filters were also used to estimate the derivatives of the data.

In order to illustrate key features, the methods in this chapter are tested on a sample displacement field. The data is from an experiment performed on a gel phantom containing four cylindrical stiff inclusions. This was done by the group at the Mayo Clinic in Dr. Richard Ehman's MRE lab [13]. The shear modulus of the gel is known to be 2.9 kPa in the background, and 6.4 kPa in the inclusions. Sinusoidal forcing was applied to one side of the phantom, at a forcing frequency of 100 Hz. This data set is described with greater detail in Chapter 7.

In order to remove noise and reflections before applying the AIDE technique, the data is filtered with a combination of a directional and band-pass filter. The directional component admits waves travelling from top to bottom, and the band-pass admits frequencies between 5 and 250 cycles/m. These values are chosen to be consistent with Manduca *et al.* [39]. The centre slice of the resulting shear modulus estimate is shown in Figure 4.1b. In this figure, it can be seen that the method has difficulty resolving the small inclusions. Instead of finding a region of stiffness (high shear modulus), it finds a region with a very small shear modulus.

The two parameters, Λ and $\mathcal{M}$, are on very different scales ($\Lambda \gg \mathcal{M}$), which can make the determination of $\mathcal{M}$ ill-conditioned. To remove this problem, Oliphant proposed that since the wavelength of compressional waves is much larger than that of the shear waves, its variation can be ignored [53]. This is the 'slow variation of longitudinal waves' assumption mentioned in Section 3.1.8. Combining this with the local homogeneity assumption, the equations become

$$-\omega^2 \mathbf{U} = \mathcal{M} \nabla \cdot \left[\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right], \tag{4.2}$$

which is an overdetermined system for a single unknown. Writing the system in matrix form,

$$\underbrace{\begin{bmatrix} U_{1,ii} + U_{i,i1} \\ U_{2,ii} + U_{i,i2} \\ U_{3,ii} + U_{i,i3} \end{bmatrix}}_{=A} \mathcal{M} = -\omega^2 \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix},$$

a least-squares solution can be obtained:

$$\mathcal{M} = -\omega^2 \left(A^{\mathrm{H}} A \right)^{-1} A^{\mathrm{H}} \mathbf{U}.$$

This method was applied to the test data set, and results are shown in Figure 4.1c. The two estimates, one with Λ and the other without, are identical. This suggests that, at least for the experimental data, the removal of the longitudinal term is justified.

The alternative to simply ignoring the longitudinal component is to apply the curl operator to both
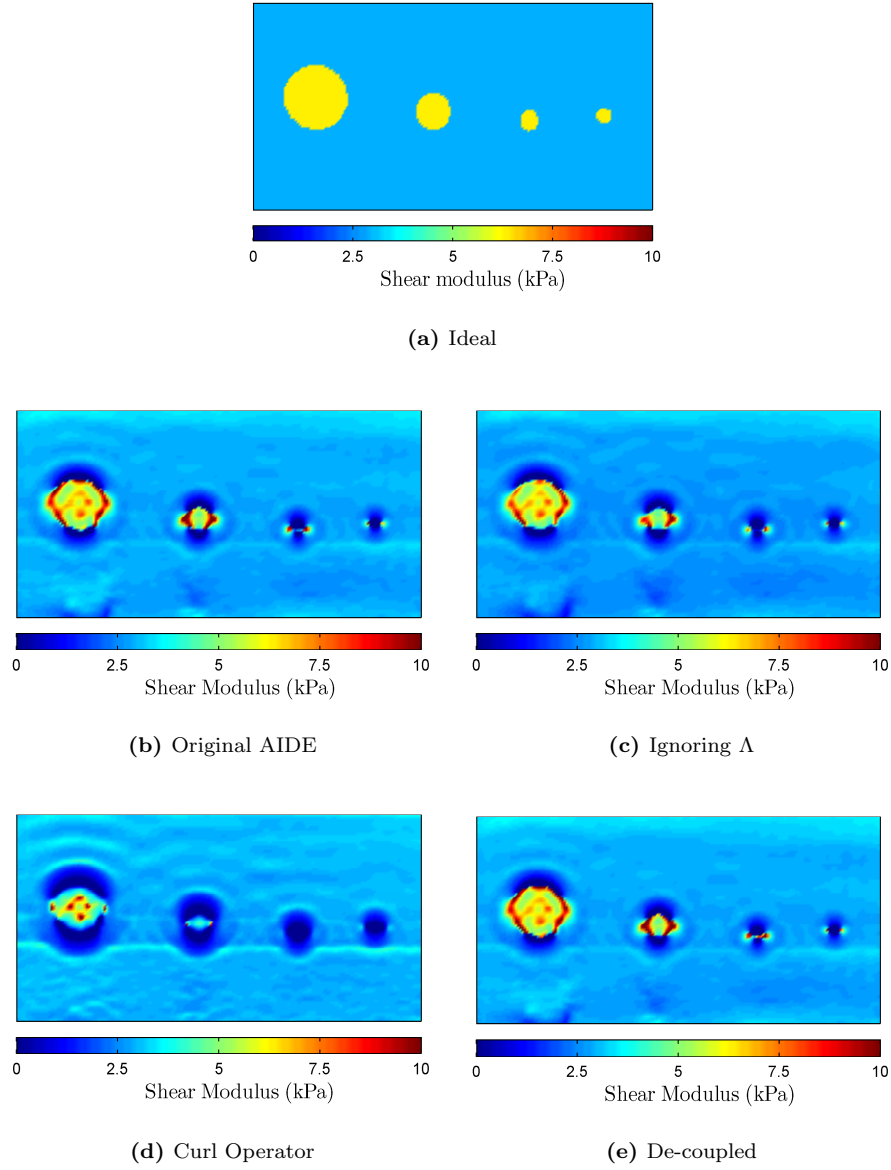
(a) Ideal

(b) Original AIDE

(c) Ignoring Λ

(d) Curl Operator

(e) De-coupled

Figure 4.1: Shear modulus estimates from the original AIDE techniques on the filtered gel phantom data set.

sides of Equation (4.1):

$$-\omega^2 \nabla \times \mathbf{U} = \mathcal{M} \, \nabla \times \nabla^2 \mathbf{U}$$

Although this requires estimating further derivatives of the displacement field, it eliminates the need to make any assumptions about the longitudinal component. Again, a new matrix of coefficients can

be constructed, and the least-squares solution obtained. The result of this estimate for the test data is shown in Figure 4.1d. Applying the curl operator increases the order of the differential equation. To still approximate the differential equation by an algebraic one, higher derivatives of the shear modulus, $\mathcal{M}$, must be assumed to be zero. This assumption is more likely to be invalid. As a result, the estimator is unable to resolve the shear modulus near any of the boundaries of the inclusions. In the figure, it is clear that there are four inclusions, but the estimate suggests that these regions are less stiff than the surrounding background, when they are actually more stiff.

The previous techniques require all three polarizations of motion. As described in Section 3.2.4, several full scans of the tissue are required to obtain just a single polarization. If the shear modulus can be accurately estimated using just one, the total scanning time can be drastically reduced. For this, the equations must be decoupled, which requires both the local homogeneity assumption and the incompressibility assumption to be made, resulting in the following Helmholtz equation:

$$-\omega^2 \mathbf{U} = \mathcal{M}\,\nabla^2 \mathbf{U}. \tag{4.3}$$

Oliphant calls this the *Interior Helmholtz* (IH) problem, and developed a set of methods based on his original AIDE technique that attempt to minimize the affect of noise [49]. The results of applying the original AIDE methodology to the IH problem are shown in Figure 4.1e. The obtained shear modulus is identical to those from the original method, where Λ is estimated, and the reduced one, where Λ is ignored. This suggests that for the test phantom, the incompressibility assumption is justified.

Once derivatives of the displacements are estimated, all the methods described so far act point-wise; the value of the shear modulus at a point depends only on $\mathbf{U}$ and its derivatives at that point. This can make the methods very sensitive to noise. To demonstrate this, the AIDE technique was applied to the unfiltered data. From Figure 4.2, it is clear that the data must be pre-processed before these original AIDE techniques are used.
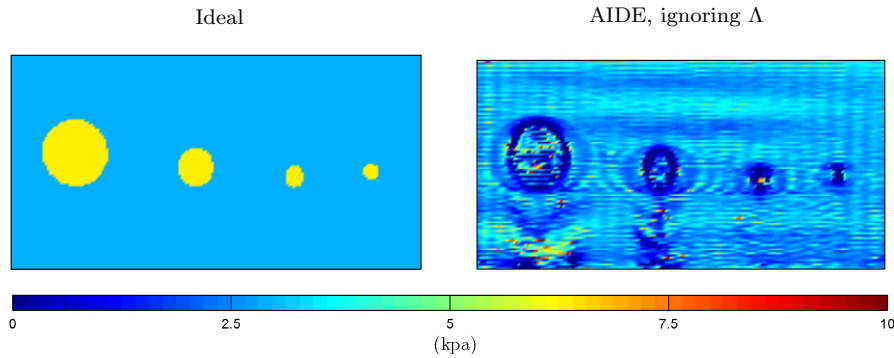

Figure 4.2: Original AIDE inversion using unfiltered data from the gel phantom experiment.

4.2   Interior Helmholtz Inversion

To reduce the problem to Interior Helmholtz inversion, the tissue must be isotropic, linearly viscoelastic, locally homogeneous, and nearly incompressible. Under these assumptions, the equations of motion

are decoupled, meaning that one polarization of motion does not affect the other two. This allows an estimate to be obtained from a single component of the displacement field.

The major issue with the original AIDE technique is its sensitivity to noise. To see the effect of noise on the shear modulus estimate, Oliphant proposed the following noise model:

$$\widehat{U} = U + \sigma W \eta, \tag{4.4}$$

where $\widehat{U}$ is a rasterized vector of the measured displacements, U is a vector of the true displacements, η is a Gaussian random vector with unit variance and uncorrelated entries, σ is a scale factor, and W is the covariance matrix of the noise. In practice, W is always taken to be diagonal, and is almost always equal to the identity matrix. The measured data can be smoothed using the techniques described in Section 3.4 in order to try recovering $\mathbf{U}$ from $\widehat{\mathbf{U}}$. However, derivatives of the displacements are still required, and as shown in the previous chapter, derivatives of even the smoothed data can be quite noisy. The only way to eliminate the amplification of noise is to eliminate the need to estimate derivatives. Oliphant accomplished this by introducing a test function, and transferring derivatives to this test function using rules of integration.

Instead of using the original system, both sides Equation (4.3) were multiplied by a smooth function, h, that has compact support:

$$-\omega^2 \, h \, U = h \, \mathcal{M} \, \nabla^2 U.$$

Such functions are often called *windows*, for when they multiply data, only information within the support of the function can be observed. The equation is integrated over the domain of h,

$$-\omega^2 \iiint\limits_{\Omega_h} h(\boldsymbol{x}) \, U(\boldsymbol{x}) \, d\boldsymbol{x} = \iiint\limits_{\Omega_h} h(\boldsymbol{x}) \, \mathcal{M}(\boldsymbol{x}) \, \nabla^2 U(\boldsymbol{x}) \, d\boldsymbol{x}.$$

Oliphant again invoked the local homogeneity assumption, assuming Ω_h is small enough that $\mathcal{M}$ can be considered roughly constant. This is a type of regularization. The resulting estimate is essentially an average over the support of h, which has a smoothing effect. This allows the shear parameter to be pulled outside of the integral:

$$-\omega^2 \iiint\limits_{\Omega_h} h(\boldsymbol{x}) \, U(\boldsymbol{x}) \, d\boldsymbol{x} = \mathcal{M}(\boldsymbol{x}) \iiint\limits_{\Omega_h} h(\boldsymbol{x}) \, \nabla^2 U(\boldsymbol{x}) \, d\boldsymbol{x}.$$

If h is chosen such that both it and its first derivatives vanish at the boundary, $\partial\Omega_h$, then the derivatives can be passed from U to the test function. This is easily derived using integration-by-parts, or by applying Green's identities. The shear modulus can then be isolated:

$$\mathcal{M}(\boldsymbol{x}) = \frac{-\omega^2 \iiint\limits_{\Omega_h} h(\boldsymbol{x}) \, U(\boldsymbol{x}) \, d\boldsymbol{x}}{\iiint\limits_{\Omega_h} U(\boldsymbol{x}) \nabla^2 h(\boldsymbol{x}) \, d\boldsymbol{x}}.$$

If there are multiple polarizations of motion that satisfy the same equation, then a best estimate of $\mathcal{M}$ can be determined using least-squares methods.

Implemented in a discrete system, the IH inversion estimate is

$$\mathcal{M} = \frac{-\omega^2 h^{\mathrm{H}} U}{(Lh)^{\mathrm{H}} U},$$

where L is the discrete Laplacian operator, h is a discretized and rasterized vector of the test function, and U is a rasterized vector of the displacements. This type of solution is known as a *weak solution*, and is only valid over the domain of h. To obtain a solution over the entire domain of measurement, a sequence of test functions is required, each with a different domain of support.

Assume h is centred at spatial position $\boldsymbol{x}_o$. Then, the weak solution at this point, $\mathcal{M}(\boldsymbol{x}_o)$, can be computed. If the test-function is then moved to be centred at the next spatial position, $\boldsymbol{x}_1$, then $\mathcal{M}(\boldsymbol{x}_1)$ can be found. In this way, a weak solution for $\mathcal{M}$ at every grid location can be determined. This moving of h by one point and integrating describes a cross-correlation. If the window is symmetric, then is equivalent to a convolution. With this methodology, Interior Helmholtz inversion becomes nothing more than a ratio of two filtered versions of the displacement component.

4.2.1   Romano AIDE

Romano *et al.* performed a similar analysis [57], but did not, however, shift the test function point-by-point, so it could not be seen as a convolution. Romano's test functions were of the following form:

$$h(\boldsymbol{x}) = \prod_{i=1}^{n} \left(1 - \left(\frac{x_i}{L_i}\right)^2\right)^2, \qquad \boldsymbol{x} \in \left\{(x_1, \ldots, x_n) \mid |x_i| < L_i,\ i = 1, \ldots, n\right\}.$$

This function satisfies the conditions that it and its first derivatives vanish at the boundary. Oliphant dubbed this the *Romano filter*, and used it in his Interior Helmholtz inversion.

The support of the window should be rather small due to the local homogeneity assumption. With few points, a discrete approximation of the window in the spatial domain can be extremely inaccurate, and derivative estimates of such a discrete approximation are even worse. However, due to the duality between the spatial and frequency domains, a function with small spatial support has a larger frequency support. Therefore, a window with very small support can be more accurately described, discretely, in the spatial-frequency domain. Discrete convolutions are typically performed in the frequency domain anyway, so this description is even more convenient. The frequency response of the Romano filter and its Laplacian are the following:

$$H(\boldsymbol{k}) = \prod_{i=1}^{n} r(k_i, L_i), \qquad r(k, L) = \begin{cases} \dfrac{16}{15}L, & k = 0 \\[2ex] \dfrac{48 - 16L^2 k^2 \sin(Lk) - 48Lk\cos(Lk)}{L^4 k^5}, & k \neq 0 \end{cases}$$

$$\nabla^2 H(\boldsymbol{k}) = -\|\boldsymbol{k}\|^2 H(\boldsymbol{k}).$$

H decays like $\boldsymbol{k}^{-3}$, and $\nabla^2 H$ like $\boldsymbol{k}^{-1}$. The size of the frequency domain should be chosen so that these functions are approximately zero at all boundaries. This can be accomplished by taking a large number of points in the discrete Fourier transform during the convolution process.

In order to test performance in the presence of noise, this method was applied to the unfiltered

data from the gel phantom experiment. Results are shown in Figure 4.3. It is apparent that the result is dependent on the support of the Romano test function. The larger the support, the smoother the solution.
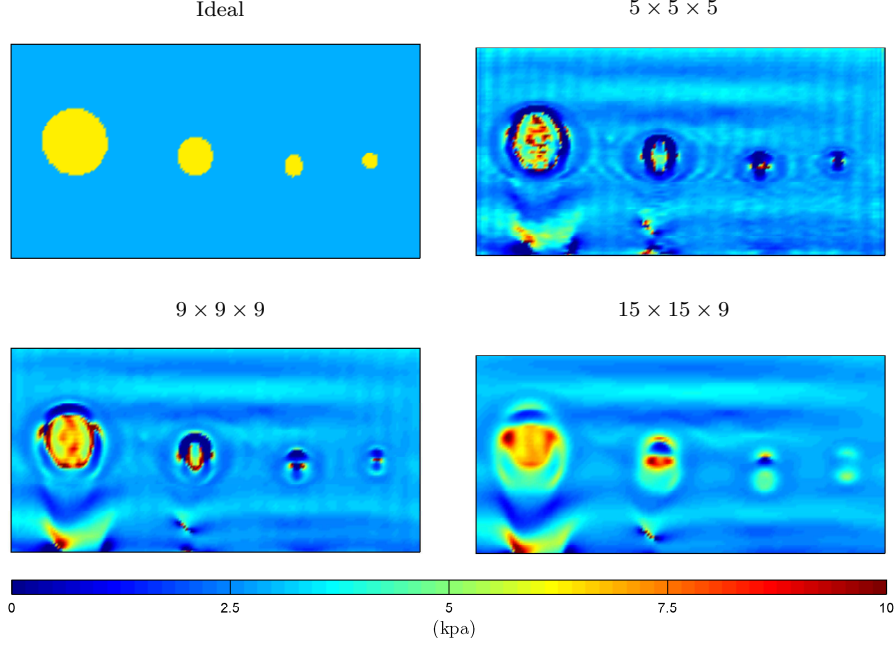

Figure 4.3: The Romano AIDE technique was applied to the unfiltered data from the gel phantom experiment, with different window sizes.

The Romano test function is not the only one that satisfies the required conditions, but it is simple and its Laplacian is easy to calculate. By passing the derivatives from the displacement to the filter, the sensitivity to noise is reduced. The choice of test function, although seemingly arbitrary, will have an impact on the estimate. In his Ph.D. thesis, Oliphant attempts to find an 'optimal' test function that minimizes the effect of noise on the inversion.

4.2.2   Optimal AIDE

In order to develop an optimal test-function, something must be known about the noise and how it affects the shear modulus reconstruction. The additive model in Equation (4.4) was used to represent the measured displacement field. With this model, an estimator and reciprocal estimator were defined:

$$\widehat{G} = \frac{(Eh)^{\textsc{h}}(\widehat{U} - \sigma W\eta)}{-(Lh)^{\textsc{h}}(\widehat{U} - \sigma W\eta)} \tag{4.5}$$

$$\hat{q} = \frac{-(Lh)^{\textsc{h}}(\widehat{U} - \sigma W\eta)}{(Eh)^{\textsc{h}}(\widehat{U} - \sigma W\eta)}. \tag{4.6}$$

The filter, h, was assumed to be extended by zeroes. The discrete operator, L, was adjusted by adding more rows than columns, so that all non-zero entries of the $\nabla^2 h$ are returned. The result is that Lh is a larger vector than h. To compensate, an extension matrix, E, is created. The product Eh returns the vector extended by zeroes so that it is the same size as Lh. These extensions implicitly impose zero Dirichlet boundary conditions when inverting the Laplacian. If the extensions are not imposed, then the Laplacian is not invertible without explicitly applying boundary conditions.

All terms in the two estimators are assumed deterministic except for the noise vector, η. The variances of the estimators are completely dependent on the variance of the noise. Under certain assumptions, Oliphant shows that if the signal-to-noise ratio (SNR) is sufficiently large, the output of the G-estimator is approximately Gaussian with the following variance:

$$\sigma_G^2 \approx \sigma^2 \frac{p_0 + p_2|G|^2 - (p_1 G + p_1^* G^*)}{|(Lh)^{\text{H}} U|^2}, \tag{4.7}$$

where

$$G = \frac{(Eh)^{\text{H}} U}{-(Lh)^{\text{H}} U},$$
$$p_0 = (Eh)^{\text{H}} WW^{\text{H}} Eh,$$
$$p_1 = -(Lh)^{\text{H}} WW^{\text{H}} Eh,$$
$$p_2 = (Lh)^{\text{H}} WW^{\text{H}} (Lh).$$

Under the same assumptions, the variance of the reciprocal estimator is also approximately Gaussian, with a variance of

$$\sigma_q^2 \approx \sigma^2 \frac{p_2 + p_0|q|^2 - (p_1 q + p_1^* q^*)}{|(Eh)^{\text{H}} U|^2}, \tag{4.8}$$

where

$$q = \frac{-(Lh)^{\text{H}} U}{(Eh)^{\text{H}} U}.$$

It was hypothesized that for each estimator, there is a single filter that minimizes the approximate output variance. These were defined as the G-optimal and q-optimal filters, respectively:

$$\tilde{h}_G = \arg\min_h \left\{ \sigma_G^2 \right\},$$
$$\tilde{h}_q = \arg\min_h \left\{ \sigma_q^2 \right\}.$$

These filters were shown to be solutions of the two fixed point equations:

$$A\left(\tilde{h}_G, U \right) \tilde{h}_G = -L^{\text{H}} U,$$
$$A\left(\tilde{h}_q, U \right) \tilde{h}_q = E^{\text{H}} U,$$

where

$$A(h,U) = L^{\mathrm{H}}WW^{\mathrm{H}}L + |q|^2 E^{\mathrm{H}}WW^{\mathrm{H}}E + q^{\mathrm{H}}L^{\mathrm{H}}WW^{\mathrm{H}}E + qE^{\mathrm{H}}WW^{\mathrm{H}}L + \frac{p_2}{r_0}E^{\mathrm{H}}UU^{\mathrm{H}}E$$

$$+ \frac{p_0}{r_0}L^{\mathrm{H}}UU^{\mathrm{H}}L + \frac{p_1}{r_0}E^{\mathrm{H}}UU^{\mathrm{H}}L + \frac{p_1^*}{r_0}L^{\mathrm{H}}UU^{\mathrm{H}}E,$$

$$r_0 = |(Eh)^{\mathrm{H}}U|^2 .$$

This can be solved iteratively,

$$h_G^{(k)} = -A\left(h_G^{(k-1)}, U\right)^{-1} L^{\mathrm{H}}U,$$

$$h_q^{(k)} = A\left(h_q^{(k-1)}, U\right)^{-1} E^{\mathrm{H}}U,$$

where k is the iteration number. Note that h depends on the noise-free data, U. Unfortunately, U is unknown; only the noisy version, $\widehat{U}$, is known. Also, since the filter depends on the data within the window, it will have to be recomputed at every single grid point. In practice, a smoothed version of the data must be used to estimate these optimal filters.
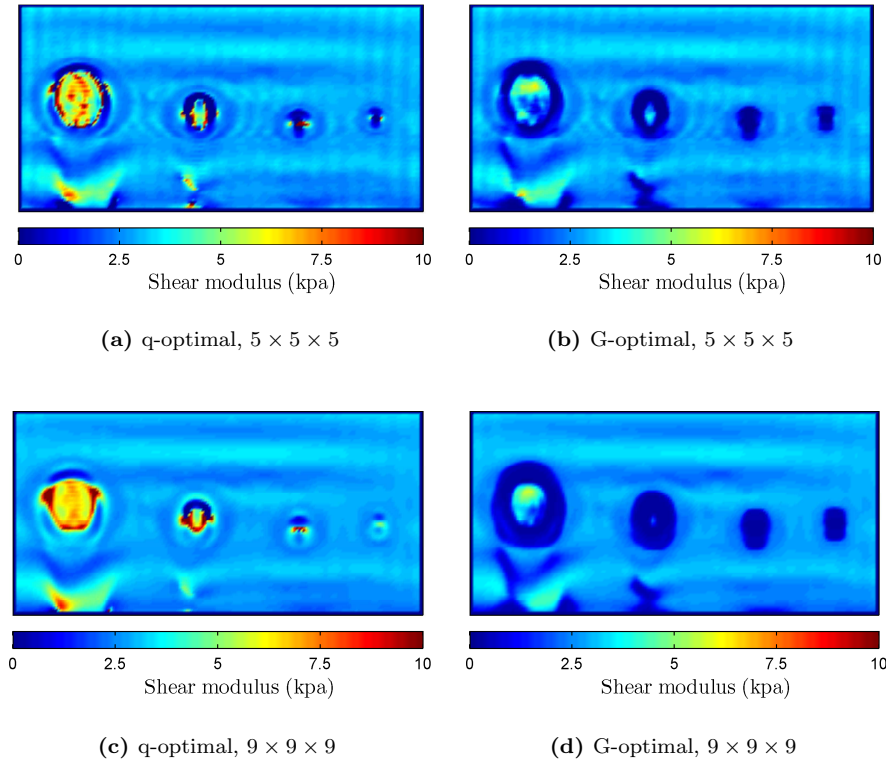

(a) q-optimal, $5 \times 5 \times 5$

(b) G-optimal, $5 \times 5 \times 5$

(c) q-optimal, $9 \times 9 \times 9$

(d) G-optimal, $9 \times 9 \times 9$

Figure 4.4: Oliphant's optimal inversion of the Interior Helmholtz equation, using unfiltered data from the gel phantom experiment.

The major flaw with this approach is that the derivation only tries to minimize the approximate variances of the estimators. It ignores the conditions that must be placed on the filter, such as being continuously differentiable, and vanishing smoothly at the boundary. In an attempt to force these conditions, L was constructed so that zero Dirichlet boundary conditions are implicitly imposed. This does not guarantee that the function goes to zero *smoothly*, however. As a result, while these methods are guaranteed to be the least-affected by noise, they are not guaranteed to be the most accurate.

Oliphant found that the q-optimal filter performed better than the G-optimal. The same results are apparent in Figure 4.4, where the optimal filters are applied to the phantom data. Oliphant also claimed that a good estimate of the q-optimal filter is $h_q^{(1)} = U$. With this initial guess, the iterative procedure converged after two or three iterations.

4.2.3 Approximately Optimal AIDE

Iteratively solving for a new optimal filter at each grid point can be cumbersome. Instead, an approximately optimal set of filter coefficients can be obtained by considering only the first, data-independent term in the fixed-point operator:

$$A(h, U) \approx L^{\text{H}} W W^{\text{H}} L$$

$$\implies \quad \tilde{h}_G \approx - \left(L^{\text{H}} W W^{\text{H}} L \right)^{-1} L^{\text{H}} U,$$

$$\tilde{h}_q \approx \left(L^{\text{H}} W W^{\text{H}} L \right)^{-1} E^{\text{H}} U.$$

The noise-free displacement is taken to be a smoothed version of $\widehat{U}$. The justification for this approximation of A comes from practical observations. By computing the individual terms in the expression for A, the first term was found to be an order of magnitude larger than the others. Since this first term is unrelated to U or h, this approximation is convenient. It also explains why the iterative procedure for determining the optimal filter coefficients converges so quickly. The method that uses $\tilde{h}_G$ was called the *Approximately-Optimal Pseudo-Inverse* (AOPI) approach, and $\tilde{h}_q$ the *Approximately-Optimal Smoothed Matched-Data* (AOSMD) approach.

The result of these methods applied to the unfiltered data set is shown in Figure 4.5. Comparing the estimates to the true optimal ones in Figure 4.4, it can be seen that there is little difference, particularly between the q-optimal and AOSMD. Because of the good agreement, only the approximately optimal inversions will henceforth be considered, since they are much faster.

4.2.4 Matched Filter AIDE

The approximate variance of the q-estimator, given in Equation (4.8), will decrease if either the numerator decreases or the denominator increases. Instead of minimizing the entire expressions, Oliphant suggested maximizing the denominator. As long as the numerator remains sufficiently small, this will have the effect of reducing the variance.

The denominator for σ_q^2 is $|h^{\text{H}} E^{\text{H}} U|$. Using a Lagrange multiplier, this can be maximized subject to
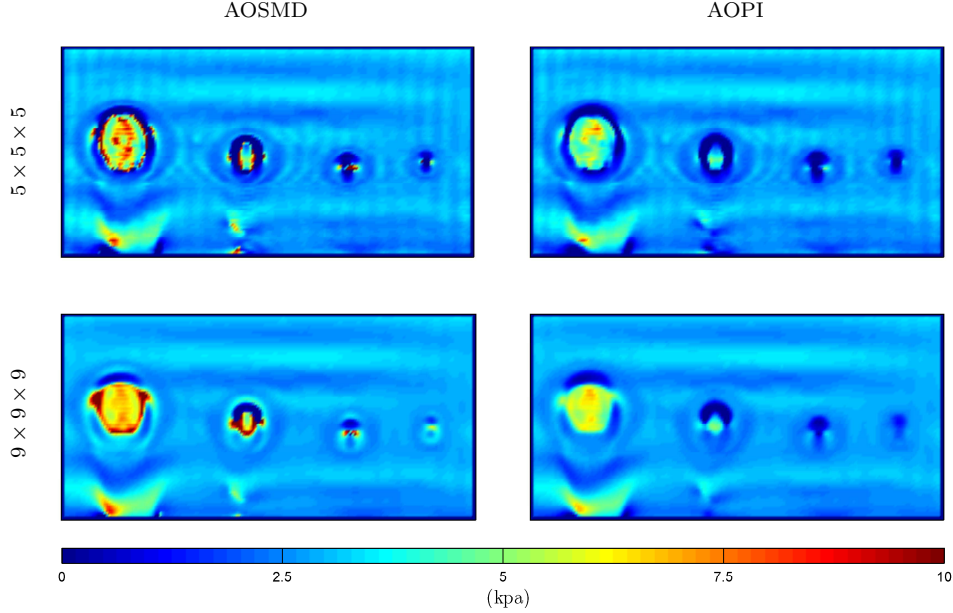
Figure 4.5: Oliphant's Approximately-optimal inversion of the Interior Helmholtz equation, using unfiltered data from the gel phantom experiment.

the constraint $\|h\| = c > 0$,

$$f(h, \lambda) = (h)^{\text{H}} E^{\text{H}} U U^{\text{H}} E h + \lambda \left(c - (h)^{\text{H}} h \right)$$

$$\implies \quad \frac{\partial f}{\partial h} = 2 E^{\text{H}} U U^{\text{H}} E h - 2\lambda h$$

$$\implies \quad \lambda h = E^{\text{H}} U U^{\text{H}} E h.$$

This is an eigenvalue problem. Since $E^{\text{H}} U$ is a column vector, $E^{\text{H}} U U^{\text{H}} E$ has a rank of one, and its image is spanned by $E^{\text{H}} U$. If h is an eigenvector with eigenvalue $\lambda \neq 0$, it must be that $h = \alpha E^{\text{H}} U$, where $\alpha \in \mathbb{R}$. Therefore, the filter that maximizes the denominator of σ_q^2 is given by $h_q = E^{\text{H}} U$. This is named the 'matched' filter. Again, since U is not known, the smoothed version must be used to estimate the optimal filter. Oliphant calls this method the Smoothed and Windowed Matched Data (SWMD) technique.

4.2.5 Modulated AIDE

In order for the q- and G-estimators to be valid, the filter and its first derivatives must go to zero at the boundary of their support. There are a plethora of possible filters that satisfy this requirement. One that was already mentioned is the Romano test function. Most of these filters, however, have a frequency spectrum that is centered about the spatial frequency $\boldsymbol{k} = 0$. If this filter is 'modulated' by multiplying by a signal, s, its frequency spectrum is shifted to centre around the frequency content of s. Therefore, if these filters are multiplied by the smoothed version of the data, the result will be tapering

filters with a frequency response that is centred around the frequency spectrum of the displacements. The modulated filters can be written as

$$h = \Phi U,$$

where Φ is a diagonal matrix with entries corresponding to a rasterized tapering window. The new filter is guaranteed to satisfy the necessary conditions, but has a better response in the frequency domain than the window alone. This can be seen as a compromise between minimizing the variation of σ_q^2 and satisfying the tapering conditions.

The results of the SWMD (Matched Filter) and Modulated AIDE techniques are compared in Figure 4.6. Notice how the modulated AIDE results are much like those from the AOSMD method.
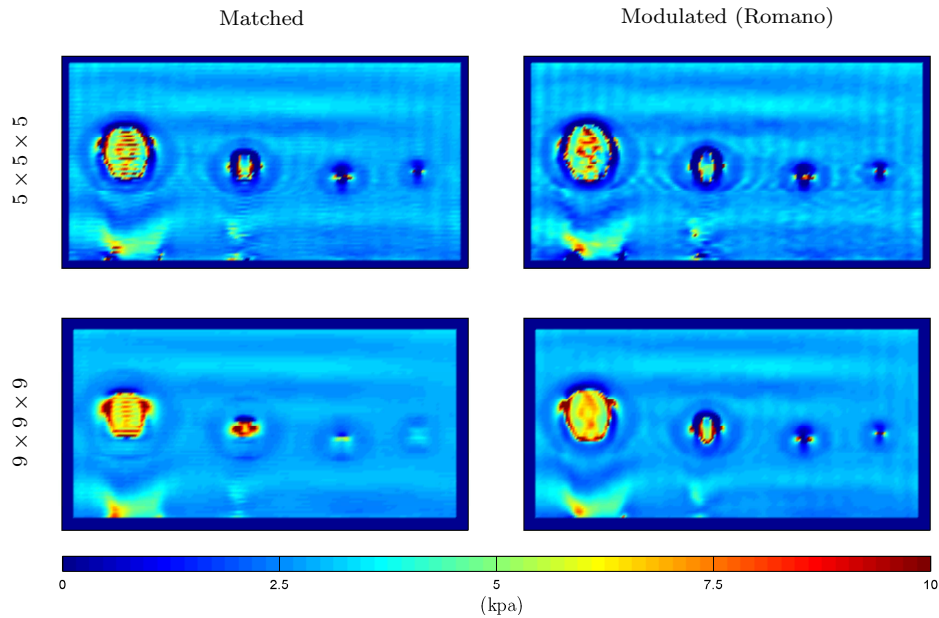

Figure 4.6: Oliphant's Matched Filter AIDE and Modulated AIDE techniques to invert the Interior Helmholtz equation, applied to the unfiltered gel phantom data.

4.2.6 Discussion

In Oliphant's work [49], the optimal filter that minimized the variance in the reciprocal estimator, q, was found to be approximately equal to the data, U. Ironically, the motivation for implementing these filters was to eliminate the need to estimate derivatives of the data; instead, the derivatives are passed onto a filter. If the filter is the data, there is no gain in this respect. In fact, taking $h = U$, the estimator

for $\mathcal{M}$ is given by

$$\begin{aligned} \mathcal{M} &= -\omega^2 \frac{U^{\text{H}}U}{(LU)^{\text{H}}U} \\ &= -\omega^2 \left(\frac{U^{\text{H}}U}{U^{\text{H}}LU} \right)^{\text{H}}. \end{aligned} \qquad (4.9)$$

Since the filter does not taper smoothly, derivatives cannot be passed form the displacement to the filter without adding boundary terms. If the original estimator is written without transferring derivatives,

$$\mathcal{M} = -\omega^2 \frac{U^{\text{H}}U}{U^{\text{H}}LU}. \qquad (4.10)$$

Comparing the two equations, it can be seen that one estimate is simply the complex conjugate of the other. Even though the filter does not taper at the edges, the only effect in Oliphant's matched filter estimate is a sign change in the complex component of the shear modulus. If one were to only examine the real component, this discrepancy would not be noticed. This may explain why the q-optimal filter seems to perform better than the G-optimal one.
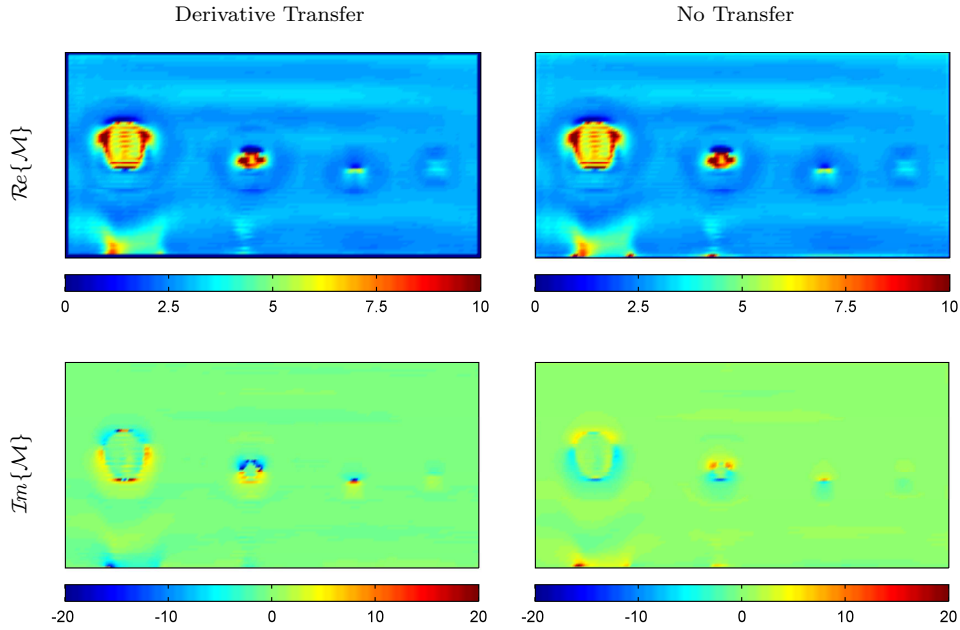

Figure 4.7: Comparison of shear modulus estimates from Matched Filter AIDE, with and without transferring the derivatives onto the filter. The window size is $9 \times 9 \times 9$.

To gain more insight into the q-optimal and G-optimal filters, reconsider the following estimator:

$$G = \frac{(Eh)^{\text{H}}(U + \sigma W\eta)}{-(Lh)^{\text{H}}(U + \sigma W\eta)}.$$

The effect of noise on the output is dependent on noise in the numerator and noise in the denominator. To minimize the total effect of noise on the estimate, the signal-to-noise ratios of both the numerator and denominator should be maximized. A total SNR can be defined of as the product:

$$\mathrm{SNR_{total}} = (\mathrm{SNR}_N)(\mathrm{SNR}_D),$$

where SNR_N and SNR_D are the signal-to-noise ratios of the numerator and denominator of G, respectively. Thus, the optimal filter is one that maximizes:

$$f(h) = \frac{1}{\sigma^4} \frac{h^\mathrm{H} E^\mathrm{H} U U^\mathrm{H} E h h^\mathrm{H} L^\mathrm{H} U U^\mathrm{H} L h}{h^\mathrm{H} E^\mathrm{H} W W^\mathrm{H} E h h^\mathrm{H} L^\mathrm{H} W W^\mathrm{H} L h}.$$

Since this is a fourth-order expression, it can be difficult to maximize. Suppose, instead, one chooses to maximize either the SNR_N or SNR_D. By the results of Section 3.4.1.5, the maximizing filters are the matched filters,

$$Eh_N = (WW^\mathrm{H})^{-1} U, \qquad\qquad Lh_D = (WW^\mathrm{H})^{-1} U,$$
$$\implies \quad h_N = E^\mathrm{H} (WW^\mathrm{H})^{-1} U, \qquad \implies \quad h_D = (L^\mathrm{H} WW^\mathrm{H} L)^{-1} L^\mathrm{H} U.$$

Notice that for $W = I$, h_N is Oliphant's matched filter, which approximately minimizes the variance of the q-estimator, and h_D is exactly the AOPI filter, which approximately minimizes the variance of the G-estimator. For these derivations, no assumptions about the distribution of the estimator itself were made. The only assumption is that the original noise in the measurements is Gaussian. It seems that the q-optimal filter simply attempts to minimize the SNR of the numerator of the shear modulus estimate, and the G-optimal filter attempts to minimize the SNR of the denominator.

Since the optimal filters are related to the data anyway, there is no reason to transfer derivatives from the data to the filter. This removes the need for the filter to taper. Without transferring the derivatives, the numerator and denominator matched filters of the estimator are the following:

$$Eh_N = (WW^\mathrm{H})^{-1} U, \qquad\qquad L^\mathrm{H} Eh_D = E^\mathrm{H} (WW^\mathrm{H})^{-1} U,$$
$$\implies \quad h_N = E^\mathrm{H} (WW^\mathrm{H})^{-1} U, \qquad \implies \quad h_D = (L^\mathrm{H} E)^{-1} E^\mathrm{H} (WW^\mathrm{H})^{-1} U.$$

The matrix $L^\mathrm{H} E$ is a square invertible matrix if L is constructed in the way described in this chapter. These filters are compared to the corresponding matched and AOPI filters of Oliphant's method in Figure 4.8. The SNR_N-minimizing estimate is identical to Oliphant's matched filter estimate, although the imaginary component has the opposite sign. The SNR_D-minimizing method, on the other hand, better resolves the small inclusions than Oliphant's AOPI.

In practice, one should avoid filters that include inverse matrices of discrete differential operators; differentiation is not invertible without enforcing some form of boundary condition. In this case, zero Dirichlet conditions are implicitly imposed by the construction of L. If L acts on data that is simply truncated, then the resulting derivatives are not continuous, which may introduce artifacts in the shear modulus solution.
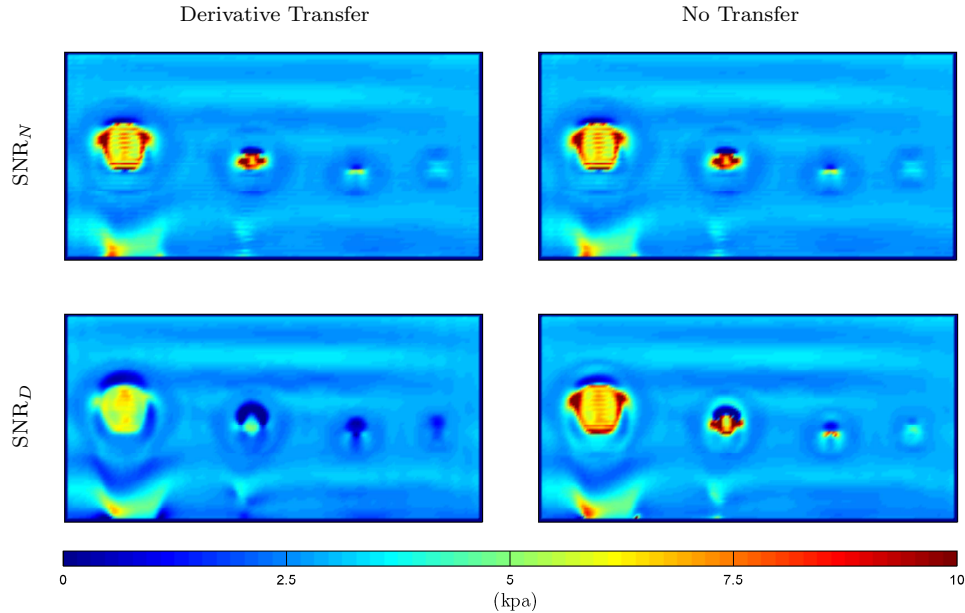
Figure 4.8: Interior Helmholtz inversion using filters that maximize the partial SNR of the estimator.

4.3 Summary

The original Algebraic Inversion of the Differential Equation technique is simple: assume $\mathcal{M}$ and Λ are locally constant (local homogeneity), pull them outside of the derivative operator, then solve the resulting overdetermined algebraic system by least-squares. The method requires all first and second order derivatives of the displacement field, which can be estimated using Savitsky-Golay filters. The system can be simplified by ignoring the longitudinal component or assuming incompressibility. The combination of both these assumptions reduces the viscoelastic wave equation to a simple Helmholtz one. Oliphant calls this simplified problem Interior Helmholtz inversion.

The AIDE method can be very sensitive to noise. This is partially because taking derivatives of the noisy data amplifies the noise, and partially because the inversion is a pointwise operation. Therefore, to reduce the sensitivity to noise, an inversion method should avoid taking derivatives of the noisy data, and/or use many points to estimate the shear modulus at one position. This was accomplished by multiplying the Helmholtz equation by a test function, integrating over the support of the function, and using rules of integration to pass derivatives off the data. By passing derivatives to the test function, fewer derivatives of the noisy data are required. Also, by integrating over the support of the test function, noise is averaged over the domain. Both of these features make this Interior Helmholtz inversion much less sensitive to noise.

The only condition on the test function is that it must go to zero smoothly at the boundary of its support. There are infinitely many test functions that satisfy this. Oliphant proposed that there is an optimal one that minimizes the effect of noise on the estimate. To find this, an additive noise model for the data was adopted, and the effect of this type of noise on the shear modulus estimate was

investigated. With a large enough SNR, the output of the shear modulus estimator can be modelled as a Gaussian random variable, with a variance dependent on the noise in the data. Two optimal filters were constructed: a G-optimal filter, that minimizes the variance of the estimator, and a q-optimal, that minimizes the variance of the reciprocal of the estimator. By experiment, it was found that the q-optimal filter has better performance.

The optimal filters need to be determined iteratively at every point in the domain, which can lead to long computation times. For this reason, approximations of the optimal filters were found. This led to the development of the AOPI and AOSMD techniques, which are the approximately-optimal versions of the G- and q-optimal methods, respectively. The q-optimal filter can also be well-approximated by the data, itself. This matched filter, however, does not necessarily taper to zero at the boundaries. To rectify this, the filter can be multiplied by another tapering function, such as the Romano filter.

The approximately optimal filters were found to be linearly related to the data. Part of the motivation for introducing the test function was to avoid taking derivatives of the data. Instead, derivatives are passed onto the test function. Since the optimal filters are linear functions of the data, differentiation of the data is still required. This defeats the purpose of the transferring. In fact, derivatives of the displacement field are better approximated when computed over the entire domain, rather than only over the support of a test function. Therefore, it is proposed that the derivatives should not be transferred. This eliminates the need for the test function to taper smoothly, so the matched filter can be used as-is.

Using a simple SNR definition of the shear modulus estimator, it is shown that Oliphant's approximately G- and q-optimal filters simply maximize the SNR of the denominator and numerator, respectively. Two new filters were proposed that minimize the denominator and numerator SNRs of the estimator, without transferring derivatives. The results are comparable to Oliphant's optimal ones.

The methods described in this chapter are the basis for the novel techniques that are derived next. The first new method is very similar to the Interior Helmholtz inversion, where a convolution is introduced in order to remove derivatives from the displacement field, and to average out noise. The second method is more similar to the original AIDE technique, where the system is converted to an algebraic one, and solved using algebraic techniques.

Direct Inversion by Green's Functions

In the direct methods of the previous chapter, filters were used to eliminate the need to differentiate noisy data. This created an integral equation for the wanted parameter, $\mathcal{M}$. In order to solve for this parameter, it was assumed constant over the support of the filter. With this assumption, the parameter was pulled outside of the integral, and an estimate was constructed through a simple ratio of filtered displacement data.

One of the aims of this dissertation is to remove the need for the local homogeneity assumption. If a special filter could be constructed that does not require $\mathcal{M}$ to be constant over its domain, then this might allow for a more accurate shear modulus estimate in regions where the shear modulus does vary.

5.1   Green's Functions

Green's functions are often used in solving inhomogeneous differential equations. If one exists, then a differential equation of the form

$$L(f)(\boldsymbol{x}) = h(\boldsymbol{x}),$$

can be solved by a simple convolution:

$$f = \mathcal{G} * h,$$

where $\mathcal{G}$ is the Green's function for the operator L, and is dependent on boundary conditions.

Green's functions do not always exist. Unfortunately, the divergence operator, which appears in the general equations of motion for a viscoelastic continuum, is an example. Instead, the Green's function for the Laplacian operator is considered. In $\mathbb{R}^3$, this has the form:

$$\mathcal{G}(\boldsymbol{x}) = -\frac{1}{4\pi\|\boldsymbol{x}\|}.$$

The derivation of this Green's function, as well as an introduction to the basic theory, is presented in Appendix A. Convolution with $\mathcal{G}$ will invert a Laplacian, provided that the function decays rapidly enough at the extremes so that the convolution is well-defined:

$$\mathcal{G} * \nabla^2 f = f.$$

If f is discontinuous, then the derivatives must be determined in the distributional, or weak, sense. In this way, the Green's function can be used to 'undo' derivatives, without having to assume the function is roughly constant over the support of the Green's function.

5.1.1   Applying Green's Functions to MRE

In order to apply the Green's function, a relation must exist in which the desired function is acted on by a Laplacian operator. Approximations are made in the equations of motion to introduce one. Begin by ignoring the longitudinal component of the elastic wave,

$$\omega^2 \mathbf{U} + \nabla \cdot (\mathcal{M} \nabla \mathbf{U}) + \nabla \cdot \left(\mathcal{M} \nabla^{\mathrm{T}} \mathbf{U}\right) = 0.$$

If $\mathcal{M}$ is roughly locally constant (local homogeneity), and if the divergence of the displacement field is small, then the first two terms dominate the balance of this expression. Under these conditions, a good approximation to the equations of motion is the following:

$$-\omega^2 \mathbf{U} = \nabla \cdot (\mathcal{M} \nabla \mathbf{U}). \tag{5.1}$$

Like the Interior Helmholtz equation, this involves decoupled motions, allowing an estimate to be obtained from a single polarization. The true Green's function for the system in Equation (5.1) depends on the data, $\mathbf{U}$, so would need to be solved numerically. To avoid this, an analytic data-independent Green's function can be obtained if the original local homogeneity assumption is made, and $\mathcal{M}$ brought *inside* the gradient operator,

$$-\omega^2 \mathbf{U} = \nabla^2 (\mathcal{M} \mathbf{U}). \tag{5.2}$$

This way, a Laplacian is acting on an expression involving the wanted parameter, $\mathcal{M}$. The difference between this and the Interior Helmholtz equation is the placement of $\mathcal{M}$. If it is assumed that $\mathcal{M}\mathbf{U}$ decays rapidly at the extremes, then the free-space Green's function for the Laplacian operator can be applied:

$$-\omega^2 \, \mathcal{G} * \mathbf{U} = \mathcal{G} * \nabla^2 (\mathcal{M} \mathbf{U})$$
$$\implies \quad -\omega^2 \, \mathcal{G} * \mathbf{U} = \mathcal{M} \, \mathbf{U}. \tag{5.3}$$

This allows for a simple estimate of the shear modulus:

$$\mathcal{M} = -\frac{\omega^2 \, \mathcal{G} * U}{U}. \tag{5.4}$$

The numerical issues that arise when performing this convolution on a discrete field are discussed in Appendix A. This expression for the shear modulus involves no derivatives whatsoever. However, it is based on the assumption that $\mathcal{M}\mathbf{U}$ is a continuous and decaying field, defined on all of $\mathbb{R}^3$. This may not be valid. In practice, displacements are only known on a finite domain, and they certainly do not taper to zero within this domain. There are two methods to compensate for this: extend the data in such a way that the appropriate conditions do apply, or formally incorporate data windows into the equations of motion.

5.1.2   Data Extensions

The measured data only exists in some finite region, Ω. Differentiation is a local operator, so no knowledge of displacements outside Ω is required for the equations of motion to hold. Introducing a convolution changes this local nature. The output of a convolution, $y = h * u$, depends on all data, u, falling within the support of the filter, h. For practical applications, an effective support can defined as follows:

$$\text{eff supp}(h, \epsilon) = \Big\{ (\boldsymbol{x}, t) \mid \|h(\boldsymbol{x}, t)\| > \epsilon \max\left(\|h(\boldsymbol{x}, t)\|\right) \Big\}, \tag{5.5}$$

where ϵ is a given tolerance level. With a tolerance of 2%, the Green's filter, for the resolution considered in this dissertation, has an effective radius of support of 40 points in the x and y directions, and 11 points in the z direction. At this size, the Green's function is almost as large as the domain of measurement, which has dimensions $188 \times 98 \times 14$. This means that, almost everywhere, outside data is required to compute the shear modulus. Since the outside data is not known, it must be invented.

The simplest way to extend the data is to assume it is zero outside the domain of measurement. With this zero-extension, the field is defined on all $\mathbb{R}^3$:

$$\widetilde{\mathbf{U}}(\boldsymbol{x}) = \begin{cases} \mathbf{U}(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

In the interior region, $\widetilde{\mathbf{U}} = \mathbf{U}$, so the full equations of motion are expected to hold,

$$-\omega^2 \widetilde{\mathbf{U}} = \nabla \cdot \left[\mathcal{M} \left(\nabla \widetilde{\mathbf{U}} + \nabla^{\mathrm{T}} \widetilde{\mathbf{U}} \right) \right] + \nabla \left[\Lambda \nabla \cdot \widetilde{\mathbf{U}} \right], \qquad \forall \boldsymbol{x} \in \Omega. \tag{5.6}$$

Outside Ω, these equations degenerate to $0 = 0$, which always holds. This new field decays at the infinities, as required by the Green's function. However, it is not continuous at $\partial\Omega$, which strongly affects the accuracy of the solution. The estimate for the unfiltered gel phantom data is shown in Figure 5.1. This resulting estimate is unreasonable; the bulk of the modulus is found to be between -20 and 20 kPa, with very high and very low peaks near the bottom. Physically, the shear modulus should always be greater than zero. The only redeeming quality of the estimate is that it is smooth; the convolution has averaged out any noise across the entire domain.
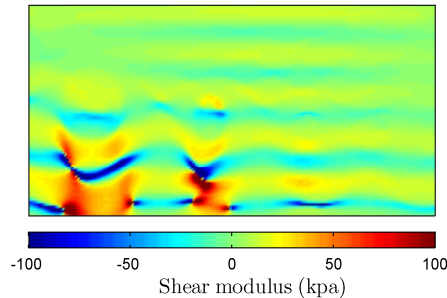

Figure 5.1: Green's function inversion using a zero extension.

To prevent the introduction of a discontinuity at $\partial\Omega$, the data can be tapered by a smooth window function, W:

$$\widetilde{\mathbf{U}}(\boldsymbol{x}) = \begin{cases} W(\boldsymbol{x})\,\mathbf{U}(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

It is assumed that there is an interior region, Ω_1, over which $W(\boldsymbol{x}) = 1$, and that W goes to zero smoothly at the boundaries. For $\boldsymbol{x} \in \Omega_1$, the equations of motion are unchanged, so are expected to hold. For $\boldsymbol{x} \in \Omega\backslash\Omega_1$, it is assumed that there exists modified parameters for which the equations of motion hold:

$$-\omega^2\widetilde{\mathbf{U}} = \nabla\cdot\left[\tilde{\mathcal{M}}\left(\nabla\widetilde{\mathbf{U}} + \nabla^{\mathrm{T}}\widetilde{\mathbf{U}}\right)\right] + \nabla\left[\tilde{\Lambda}\,\nabla\cdot\widetilde{\mathbf{U}}\right], \qquad \forall\boldsymbol{x} \in \Omega\backslash\Omega_1. \tag{5.7}$$

In this region, $\tilde{\mathcal{M}} \neq \mathcal{M}$; it is modified to account for the tapering of the data. Thus, the determined shear modulus is only valid for points $\boldsymbol{x} \in \Omega_1$.
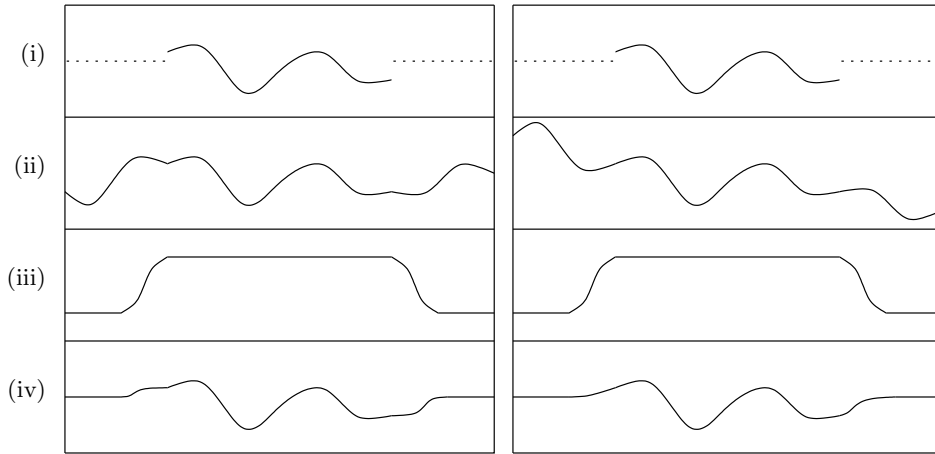

Figure 5.2: Data extension with tapering. (i): The original data. (ii) Even (left) or odd (right) extension. (iii) Tapering window. (iv) Tapered extension.

If the estimate is desired near any of the boundaries, then the data must first be continuously extended before tapering. This can be done by performing either an even or odd extension at all boundaries. A one-dimensional example of this is shown in Figure 5.2. With both even and odd extensions, $\widetilde{\mathbf{U}}$ is guaranteed to be continuous. The odd extension also has the advantage that the first derivative is continuous. Applying the Green's function to the equation with extended and tapered data results in

$$-\omega^2\,\mathcal{G} * \left(W\widetilde{U}\right) = \tilde{\mathcal{M}}W\widetilde{U}. \tag{5.8}$$

The modified shear modulus within the support of the window is estimated by

$$\tilde{\mathcal{M}} = -\frac{\omega^2 \, \mathcal{G} * \left(W\widetilde{U}\right)}{W\widetilde{U}}, \qquad \forall \boldsymbol{x} \in \Omega.$$

Results for both even and odd tapering are shown in Figure 5.3. These estimates show similar artifacts to the one with the zero extension, suggesting that simply extending and tapering the data is a naïve approach.
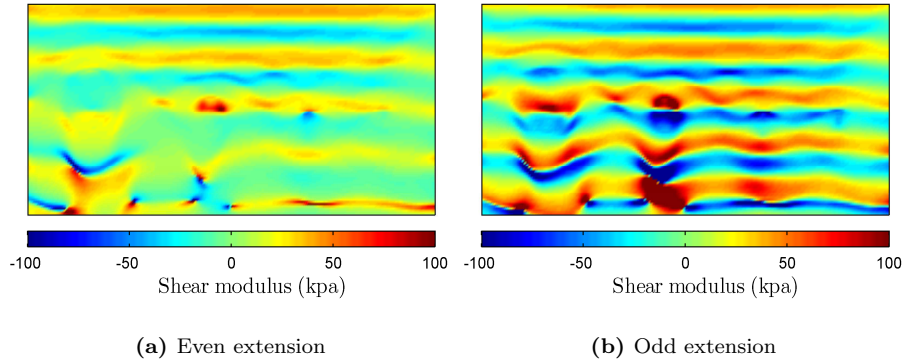

(a) Even extension   **(b)** Odd extension

Figure 5.3: Green's function inversion using even and odd tapered data extensions.

It has been asserted that convolution with $\mathcal{G}$ is an almost global operation. This means the solution at a point depends on the data almost everywhere, including the extension. The choice of extension, therefore, has a direct effect on the solution for the shear modulus. At any discontinuous boundary, and within the tapering regions, the original equations of motion no longer hold. A convolution then smears this error over the entire domain, causing large errors in the estimate, everywhere. There seems to be a missing term in Equation (5.3) to account for the finite boundary of the data. This is explored in the next section.

5.1.3   Data Windows

The alternative to extending the data is to formally introduce a window into the equations of motion. The equations, themselves, should hold, provided displacements and parameters are continuous on all $\boldsymbol{x} \in \mathbb{R}^3$. However, there is no knowledge of these continuous displacements outside of the domain of measurement. If the entire equation is multiplied by a window function whose support lies within Ω, then this truncated set of equations is still expected to hold everywhere:

$$-\omega^2 \, W\mathbf{U} = W\nabla \cdot \left[\mathcal{M}\left(\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right)\right] + W\nabla\left[\Lambda \, \nabla\cdot\mathbf{U}\right], \qquad \forall \boldsymbol{x} \in \mathbb{R}^3. \tag{5.9}$$

The window, W, belongs to the space of generalized functions (or distributions) that have compact support. Since differentiation operators are local, this truncated system of equations does not include any information outside the domain of measurement.

The same local homogeneity and dominant term assumptions are made as before, giving

$$-\omega^2 \, W\mathbf{U} = W\nabla^2 \left(\mathcal{M}\mathbf{U}\right). \tag{5.10}$$

Before applying the Green's function, the right-hand side is modified using the following vector calculus identity:

$$\nabla^2 \left(fg\right) = f\nabla^2 g + 2\nabla f \cdot \nabla g + g\nabla^2 f. \tag{5.11}$$

With this identity, the window function, W, can be brought inside the Laplacian operator,

$$-\omega^2 \, W\mathbf{U} = \nabla^2 \left(\mathcal{M}\, W\mathbf{U}\right) - 2\nabla W \cdot \nabla \left(\mathcal{M}\mathbf{U}\right) - \mathcal{M}\mathbf{U}\nabla^2 W. \tag{5.12}$$

A simple window that does not modify the measured data is a product of the discontinuous boxcar functions:

$$W_\square(x,y,z) = \left[H(x - x_l) - H(x - x_r)\right]\left[H(y - y_l) - H(y - y_r)\right]\left[H(z - z_l) - H(z - z_r)\right],$$

where H is the Heaviside function, and the domain of measurement is

$$\Omega = \left\{(x,y,z) \mid x_l \le x \le x_r, \ y_l \le y \le y_r, \ z_l \le z \le z_r\right\}.$$

Since this window is discontinuous, derivatives must be considered in the distributional sense. Applying the Green's function to Equation (5.12) results in

$$-\omega^2 \, \mathcal{G} * \left(W\mathbf{U}\right) = \mathcal{M}\, W\mathbf{U} - \underbrace{\mathcal{G} * \left(2\nabla W \cdot \nabla \left(\mathcal{M}\mathbf{U}\right) + \mathcal{M}\mathbf{U}\nabla^2 W\right)}_{\mathcal{B}}. \tag{5.13}$$

Compare this to Equation (5.8). The last term on the right-hand side, $\mathcal{B}$, is the missing one from the previous section. It accounts for the fact that the displacements are unknown outside Ω.
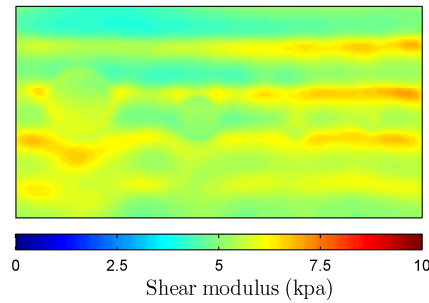

Figure 5.4: Initial estimate of shear modulus using DIG with a boxcar window.

Without further assumptions, the system (5.13) must be solved iteratively for $\mathcal{M}$. For an initial estimate, one can resort to the AIDE methodology, and assume $\mathcal{M}$ is constant over the support of the

derivatives of the window:

$$-\omega^2 \, \mathcal{G} * (W\mathbf{U}) = \mathcal{M}_1 \left[W\mathbf{U} - \mathcal{G} * \left(2\nabla W \cdot \nabla \mathbf{U} + \mathbf{U}\nabla^2 W \right) \right].$$

$$\implies \quad \mathcal{M}_1 = -\frac{\omega^2 \, \mathcal{G} * (W\mathbf{U})}{W\mathbf{U} - \mathcal{G} * (2\nabla W \cdot \nabla \mathbf{U} + \mathbf{U}\nabla^2 W)} \tag{5.14}$$

With this initial estimate, Equation (5.13), along with some iterative scheme, can be used to find the shear modulus over the domain of the window. This method will be referred to as *Direct Inversion by Green's functions* (DIG). A subscript 1 is used to denote the initial estimate. The development of an iterative scheme has proved difficult because of the global nature of the convolution. It has been left for future work.

The initial estimate, applied to the unfiltered gel phantom data, is shown in Figure 5.4. The shear modulus is at least on the correct order of magnitude, and is positive everywhere. The mean value is near the true average stiffness of the gel phantom. The homogeneity assumptions, however, smooth out the estimate so that the inclusions are missed.

5.1.3.1   Examples of windows

Window functions for the DIG method are considered in the most general sense; they belong to the set of distributions (or generalized functions) with compact support. These generalized functions act on the space of continuous functions. This means that the displacements within the window must be continuous. So far, only two windows have been considered in this dissertation: the Romano test function, and the boxcar function. In this section, properties of these two windows are reviewed, and a set of more desirable windows are derived.

Boxcar

The boxcar function is discontinuous, which means its derivatives must be taken in the sense of distributions. In the MRE application, both first and second derivatives of the window are required. To analyze these derivatives, consider the one-dimensional boxcar:

$$W_\square(x) = H(x - x_l) - H(x - x_r),$$

where H is the Heaviside function. The first derivative of this, in the sense of distributions, is

$$\frac{\partial W_\square}{\partial x}(x) = \delta(x - x_l) - \delta(x - x_r),$$

where δ is the Dirac delta function. These functions are only defined through an inner-product:

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_o)dx = f(x_o).$$

The second derivative of $W_\square$ introduces the δ' distribution:

$$\frac{\partial^2 W_\square}{\partial x^2}(x) = \delta'(x - x_l) - \delta'(x - x_r).$$

These behave in the following way:

$$\int_{-\infty}^{\infty} f(x)\delta'(x - x_o)dx = -\int_{-\infty}^{\infty} \frac{\partial f}{\partial x}(x)\delta(x - x_o)dx = -\frac{\partial f}{\partial x}(x_o).$$

With this formalism, second derivatives of the three dimensional boxcar function can be derived. Using the boxcar window in Equation (5.13), the boundary term becomes

$$\mathcal{B}(\boldsymbol{x}) = \int_{\Omega} \mathcal{G}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\alpha - x_l) - \delta(\alpha - x_r)\right] \frac{\partial f}{\partial \alpha}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} + \int_{\Omega} \mathcal{G}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\beta - y_l) - \delta(\beta - y_r)\right] \frac{\partial f}{\partial \beta}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha}$$

$$+ \int_{\Omega} \mathcal{G}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\gamma - z_l) - \delta(\gamma - z_r)\right] \frac{\partial f}{\partial \gamma}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} - \int_{\Omega} \frac{\partial \mathcal{G}}{\partial x}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\alpha - x_l) - \delta(\alpha - x_r)\right] f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha}$$

$$- \int_{\Omega} \frac{\partial \mathcal{G}}{\partial y}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\beta - y_l) - \delta(\beta - y_r)\right] f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} - \int_{\Omega} \frac{\partial \mathcal{G}}{\partial z}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\gamma - z_l) - \delta(\gamma - z_r)\right] f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha},$$

where $f = \mathcal{M}\mathbf{U}$, $\boldsymbol{\alpha} = (\alpha, \beta, \gamma)$, $\boldsymbol{x} = (x, y, z)$, and Ω is the support of the window. Notice that only first derivatives of the displacement field are required. This reduction in derivative-order improves the stability of the algorithm in the presence of noise.

Polynomial transitions

The δ and δ' distributions can be difficult to work with, numerically. To prevent these from appearing, W must belong to the space of $C^1(\mathbb{R})$ functions. In this section, simple windows are constructed using a combination of piecewise polynomial and constant functions.

Consider a single dimension. If the window is to be continuously differentiable and have support in $[-b, b]$, then it must satisfy $W(\pm b) = 0$ and $W'(\pm b) = 0$. It is easily shown that this means it must have repeated roots at $x = \pm b$. This leads to the simple window,

$$W(x) = a(b - x)^2(b + x)^2 = a\left(b^2 - x^2\right)^2, \quad x \in [-b, b],$$

where a is some scaling constant. If the centre of the window is chosen to have a value $W(0) = 1$, the result is the Romano test function.

The Romano test function only evaluates to one at a single point in the domain $[-b, b]$. If this window is used for tapering data to zero, then only a single point (the centre) is unaffected. It may be desirable to have a range of points that are unaffected. For this reason, windows are constructed to have the following form:

$$W_p(x) = \begin{cases} p(x, -a, -b), & x \in [-b, -a] \\ 1, & x \in [-a, a] \\ p(x, a, b), & x \in [a, b] \\ 0, & \text{otherwise} \end{cases},$$

where p is a transition polynomial that forces continuity. These windows have a compact support in $[-b, b]$, are symmetric, and also satisfy the condition that $W_p(x) = 1$ for $x \in [-a, a]$. For the window to be continuous, p must satisfy $p(a) = 1$ and $p(b) = 0$. For the window to have a continuous derivative, the transition polynomial must also satisfy $p'(a) = 0$ and $p'(b) = 0$.

The lowest-order polynomial that satisfies the required conditions has order three. This is clear, since p' has two roots, so integrating leads to three. For $W_p \in C^2$, p' must have repeated roots at a and b, which means p is of order five. Similarly, for $W_p \in C^3$, p' must have triple roots at a and b, so p must be a seventh-order polynomial. These transition polynomials are given by the following equations:

$$p_3(x, a, b) = \frac{(b - x)^2(3a - b - 2x)}{(a - b)^3}$$

$$p_5(x, a, b) = \frac{(b - x)^3(10a^2 - 5ab - 15ax + 3xb + b^2 + 6x^2)}{(a - b)^5}$$

$$p_7(x, a, b) = \frac{(b - x)^4(35a^3 - b^3 + 7ab^2 - 4xb^2 - 21a^2b - 10x^2b + 28abx - 20x^3 + 70ax^2 - 84a^2x)}{(a - b)^7}.$$

With these transition polynomials, smooth windows can be constructed that satisfy the desired conditions. The one-dimensional windows are plotted in Figure 5.5.
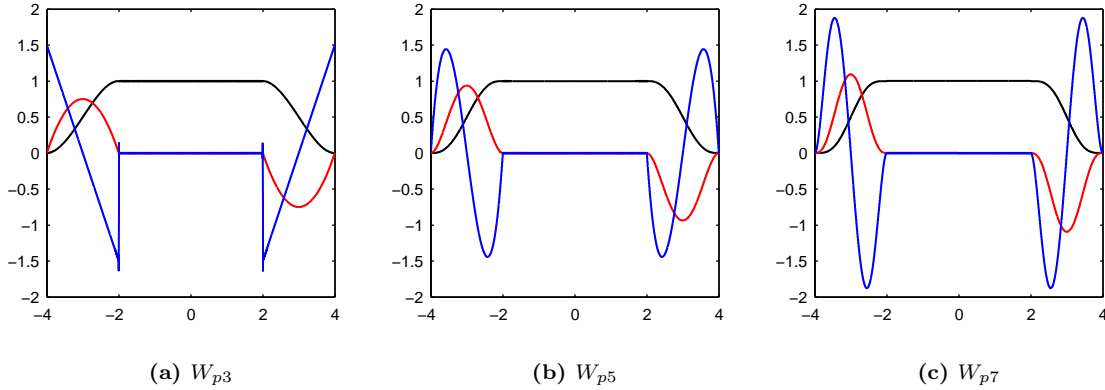

(a) W_{p3} **(b)** W_{p5} **(c)** W_{p7}

Figure 5.5: Windows with polynomial transitions. The one-dimensional window (black), its first derivative (red), and its second derivative (blue) are plotted for third, fifth, and seventh order polynomial transitions.

These windows will be used on discrete data in the application. If the support is small, relative to the resolution of the displacement field, they should be constructed in the spatial-frequency domain. This is because functions with small spatial support have a much larger frequency support. This allows for a better discrete representation of the windows. The frequency description of the one-dimensional windows are

$$W_{p3}(k) = \begin{cases} a + b, & k = 0 \\ \dfrac{24\left[\cos(ak) - \cos(bk)\right] + 12(a - b)k\left[\sin(ak) + \sin(bk)\right]}{(b - a)^3 w^4}, & k \neq 0 \end{cases}$$

$$W_{p5}(k) = \begin{cases} a+b, & k=0 \\ \dfrac{120\left[12-(a-b)^2k^2\right]\left[\cos(ak)-\cos(bk)\right]+120\left[6(a-b)k\right]\left[\sin(ak)+\sin(bk)\right]}{(b-a)^5w^6}, & k\neq 0 \end{cases}$$

$$W_{p7}(k) = \begin{cases} a+b, & k=0 \\ \dfrac{1680\left[120-12(a-b)^2k^2\right]\left[\cos(ak)-\cos(bk)\right]}{(b-a)^7w^8} \\ \quad +\dfrac{1680\left[60(a-b)k-(a-b)^3k^3\right]\left[\sin(ak)+\sin(bk)\right]}{(b-a)^7w^8}, & k\neq 0 \end{cases}$$

Higher dimensional windows can be formed by considering the product of one-dimensional ones:

$$\boldsymbol{W}_p(\boldsymbol{k}) = \prod_{i=1}^{n} W_p(k_i).$$

Derivatives are readily available through the relationship: $\frac{\partial \boldsymbol{W}}{\partial x_i}(\boldsymbol{k}) = ik_i\boldsymbol{W}(\boldsymbol{k})$. Since discrete convolutions are typically performed in the frequency domain, this frequency representation has an added advantage. These are now ready to be applied to the DIG inversion method.

5.1.4   Block Inversion

In the gel phantom test-case, it is known that the object contains four cylindrical regions of stiffness. These are not found by the DIG_1 inversion in Figure 5.4. Due to the assumptions for the initial estimate, the shear modulus is smoothed out over the domain, hiding the stiff inclusions. The assumptions may, however, be more applicable on smaller domains.

The original equations of motion are completely local; they hold pointwise, suggesting $\mathcal{M}$ might be solved in pointwise fashion. This was done in the original AIDE technique. Once the convolution is introduced, this nature is changed. The large support of the Green's function makes the system almost global. This has both advantages and disadvantages. A large advantage is that the measurement and numerical noise is smeared out over the domain, resulting in a much smoother solution. One major disadvantage is that systematic errors and large spikes in the data, particularly along boundaries and edges of inclusions, now affect the solution everywhere. When windows are introduced, the estimate of $\mathcal{M}$ at a point depends only on the data within the support of the window. This allows noise to be spread out, but also limits the impact of large outliers and systematic errors.

Working with windows allows the domain to be broken up into smaller sub-domains, on which the shear modulus can be estimated independently. This is much like Van Houten's subzone technique [68]. Working on smaller domains has computational advantages as well. Any iterative procedures, like attempting to solve Equation (5.13), can be performed much more quickly on a smaller domain. This type of problem is known as being *embarrassingly parallel* [16]. Since the inversion can be performed on different subdomains independently, the sub-problems can be solved in parallel, and the solutions merged.

Once a suitable window size is selected, the placements of these windows must be decided. One
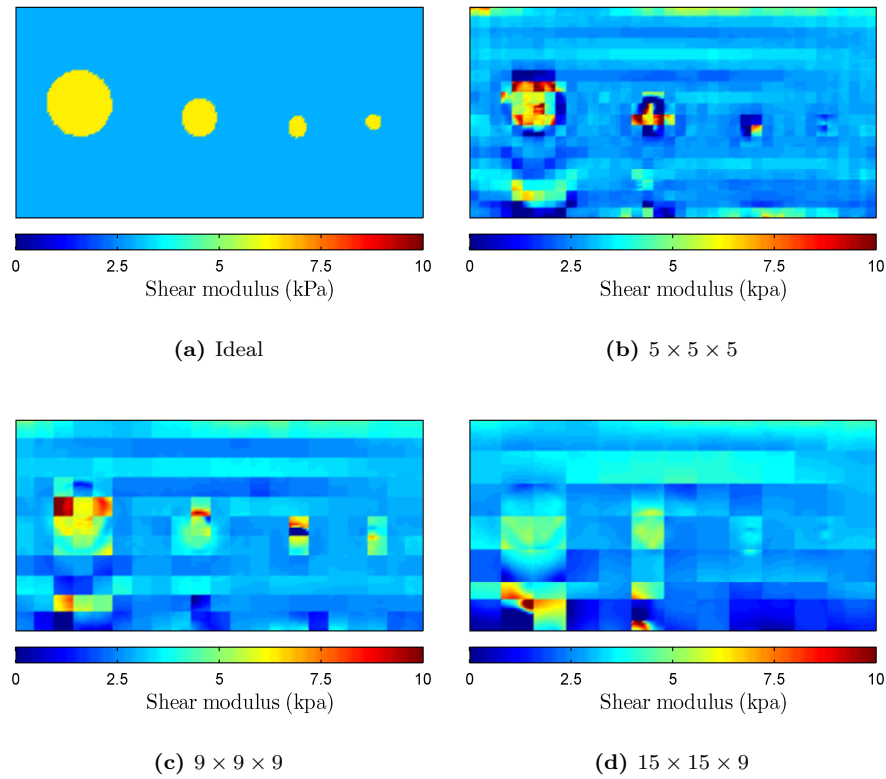
(a) Ideal

(b) $5 \times 5 \times 5$

(c) $9 \times 9 \times 9$

(d) $15 \times 15 \times 9$

Figure 5.6: DIG_1 block inversion on the unfiltered gel phantom data.

option is to consider disjoint blocks. This might be the most efficient, since an estimate at each point is calculated only once. The global estimate is then composed by patching together the windowed estimates, side-by-side. An example of such a solution is shown in Figure 5.6. With this type of construction, the global estimate tends to have block-like artifacts, showing the boundaries of the subdomains. If the subdomains are selected so that there is an overlap, the solutions in the overlapping regions can be merged smoothly. This can help create a smooth global solution.

If there is no reason to favour the estimate from one block over another, then the most appropriate merging procedure is to average the solutions in the overlap. However, this can still lead to discontinuous results. For example, if the estimate from one block revealed the shear modulus to have a value of 1 kPa, but the next block an estimate of 3 kPa, then the merged solution would discontinuously jump from 1 to 2 to 3 kPa. A better assumption might be that values closer to the centre of a window are more reliable. In this case, a weighted average should be used. These weights can be chosen to guarantee a smooth transition from block to block. The simplest is a linear averaging, as shown in Figure 5.7. In this type of merging, the boundaries of the blocks have a zero weight, while the centres have a weight of one, allowing a linear transition, e.g. the shear modulus will linearly change from 1 kPa to 3 kPa.

Examples of a block solutions with overlaps and linear merging are shown in Figure 5.8. In this figure, it is verified that the block artifacts are removed, resulting in a smooth global estimate.
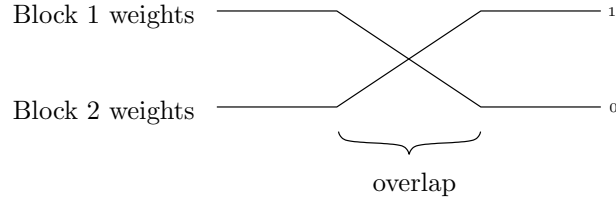
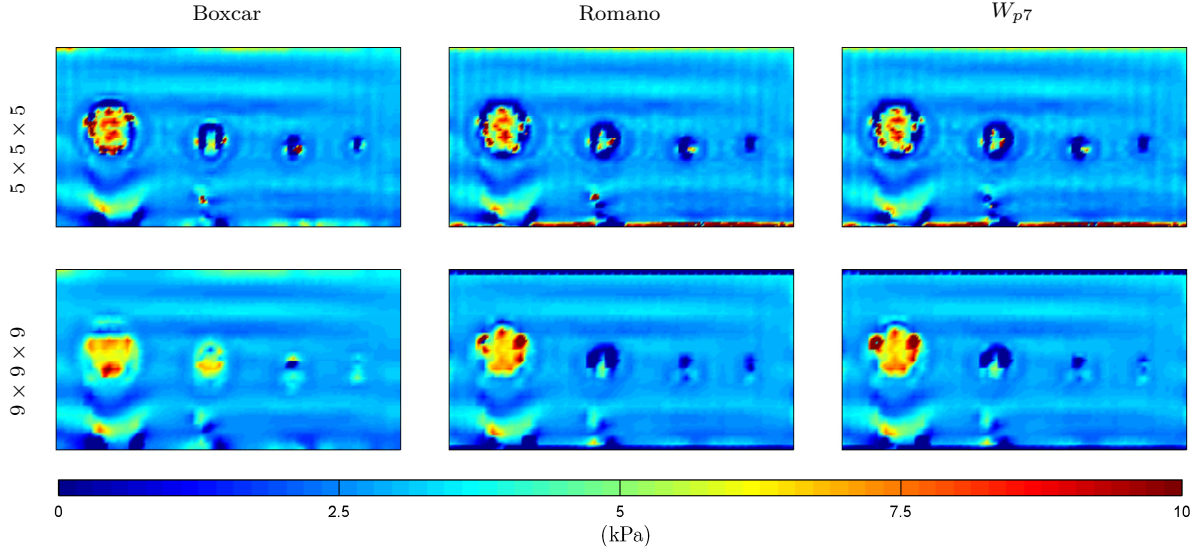**Figure 5.7:** Linear transition in overlapping region.

Figure 5.8: DIG$_1$ overlapping block inversion on unfiltered data from the gel phantom experiment. Blocks are overlapped by 50%, and merged linearly.

5.1.5   Moving Window Inversion

Consider the case when the window is centred at a point, say $\boldsymbol{x}_o$, and only the value of $\mathcal{M}(\boldsymbol{x}_o)$ is kept from the estimate. This is how Oliphant proceeded with the IH inversion. The window can then be moved to provide an estimate at the next point. This describes a convolution process. Beginning with Equation (5.10), applying the Green's function, and evaluating the equation at $\boldsymbol{x} = \boldsymbol{x}_o$, the following is obtained:

$$-\delta\left(\boldsymbol{x} - \boldsymbol{x}_o\right)\omega^2\,\mathcal{G} * \left[W(\boldsymbol{x} - \boldsymbol{x}_o)\mathbf{U}\right] = \delta\left(\boldsymbol{x} - \boldsymbol{x}_o\right)\mathcal{G} * \left[W(\boldsymbol{x} - \boldsymbol{x}_o)\nabla^2\left(\mathcal{M}\mathbf{U}\right)\right]$$
$$\implies \quad -\omega^2\left(W\mathcal{G}\right) * \mathbf{U} = \left(W\mathcal{G}\right) * \nabla^2\left(\mathcal{M}\mathbf{U}\right).$$

Here, it is assumed that the window is symmetric, which is the case for all windows mentioned. Using properties of a convolution, and expanding using the vector calculus identity in Equation (5.11), this

equation becomes

$$-\omega^2 \left(W\mathcal{G}\right) * \mathbf{U} = \nabla^2 \left(W\mathcal{G}\right) * \left(\mathcal{M}\mathbf{U}\right) \tag{5.15a}$$

$$= \mathcal{M}\mathbf{U} + 2(\nabla W \cdot \nabla \mathcal{G}) * (\mathcal{M}\mathbf{U}) + (\mathcal{G}\nabla^2 W) * (\mathcal{M}\mathbf{U}). \tag{5.15b}$$

As an initial estimate, $\mathcal{M}$ can be assumed constant over the support of the window's derivatives:

$$-\omega^2 \left(W\mathcal{G}\right) * \mathbf{U} = \mathcal{M}_1 \nabla^2 \left(W\mathcal{G}\right) * \mathbf{U}$$

$$\implies \qquad \mathcal{M}_1 = -\frac{\omega^2 \left(W\mathcal{G}\right) * \mathbf{U}}{\nabla^2 \left(W\mathcal{G}\right) * \mathbf{U}} \tag{5.16}$$

This is exactly the form used in the IH inversion, with the filter $h = W\mathcal{G}$. The estimate of $\mathcal{M}$ can then be improved iteratively with Equation (5.15). This method uses a *moving window*, so will be referred to as DIG-MW. Shear estimates from this method for the unfiltered gel phantom data are shown in Figure 5.9. Notice that there does not seem to be a strong dependence on the window type. The boxcar window produces a slightly smoother estimate. The homogeneity assumption causes the shear modulus to be averaged over the support of the window. Since the boxcar function is constant over its support, the shear modulus is evenly averaged. For the windows that taper smoothly, more weight is placed on the values that lie near the centre. This explains why the boxcar estimate is smoother than the others for a fixed window size.
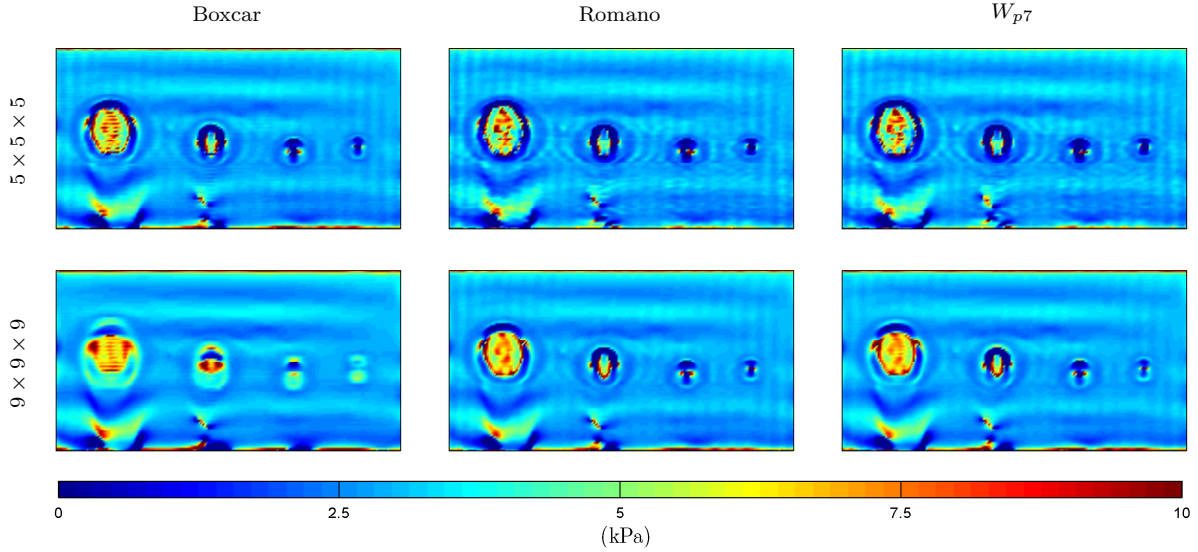

Figure 5.9: DIG_1 estimate with moving window (DIG_1-MW), applied to the unfiltered gel phantom data.

5.2 Potentials

The Green's function approach described in this chapter is very similar to Oliphant's IH inversion, but tries to remove the assumption that the elasticity is constant over the support of the filter. Without an iterative procedure, a slightly weaker homogeneity assumption is required: the elasticity must be constant over the support of the derivatives of the filter. The local homogeneity assumption was also needed to introduce the Laplacian operator. In an attempt to remove this, consider a split system:

$$-\omega^2 \mathbf{U} = \nabla^2 (\Phi \mathbf{U}) + \nabla [\Lambda \nabla \cdot \mathbf{U}] \tag{5.17a}$$

$$\nabla \cdot [\mathcal{M} (\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U})] = \nabla \cdot [\nabla (\Phi \mathbf{U})] . \tag{5.17b}$$

The intermediate function, Φ, can be thought of as a type of 'potential' for $\mathcal{M}$. This new system is equivalent to the original viscoelastic wave equation with unit density. However, Φ can be solved using the formal introduction of the window, then an estimate of the shear modulus can later be obtained using (5.17b). This latter relation has the solution:

$$\mathcal{M} (\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U}) = \nabla (\Phi \mathbf{U}) + \boldsymbol{H},$$

where $\boldsymbol{H}$ is a matrix field that satisfies $\nabla^2 \boldsymbol{H} = 0$, and is determined by the boundary conditions placed on $\mathcal{M}$. Further regularization can be imposed by forcing $\boldsymbol{H} = 0$:

$$-\omega^2 \mathbf{U} = \nabla^2 (\Phi \mathbf{U}) + \nabla [\Lambda \nabla \cdot \mathbf{U}] , \tag{5.18a}$$

$$\mathcal{M} (\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U}) = \nabla (\Phi \mathbf{U}) \tag{5.18b}$$

Solutions, $\mathcal{M}$, of system (5.18) are guaranteed to satisfy (5.17), but the converse is not true. The regularization has narrowed the space of admissible solutions. Hopefully, this did not remove the 'true' solution.

Further simplifications include ignoring the longitudinal component,

$$-\omega^2 \mathbf{U} = \nabla^2 (\Phi \mathbf{U}) , \tag{5.19a}$$

$$\mathcal{M} (\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U}) = \nabla (\Phi \mathbf{U}) , \tag{5.19b}$$

or the coupled terms,

$$-\omega^2 \mathbf{U} = \nabla^2 (\Phi \mathbf{U}) , \tag{5.20a}$$

$$\mathcal{M} \nabla \mathbf{U} = \nabla (\Phi \mathbf{U}) . \tag{5.20b}$$

Notice that by imposing the local homogeneity assumption in Equation (5.20b), Φ becomes equal to $\mathcal{M}$, and Equation (5.2) is recovered. This suggests that the introduction of this potential can be thought of as a small correction that allows for some variability in $\mathcal{M}$.

The result of solving the system in Equation (5.20) is shown in Figure 5.10. This introduction of the potential has removed some of the small background waves in the image. It also sharpens the transition of the shear modulus around the boundaries of the inclusions. Without the potential, the inclusions seem more oval-shaped, when they should be circular. Unfortunately, this method also introduces large spikes, and overestimates the values within the cylindrical inclusions.
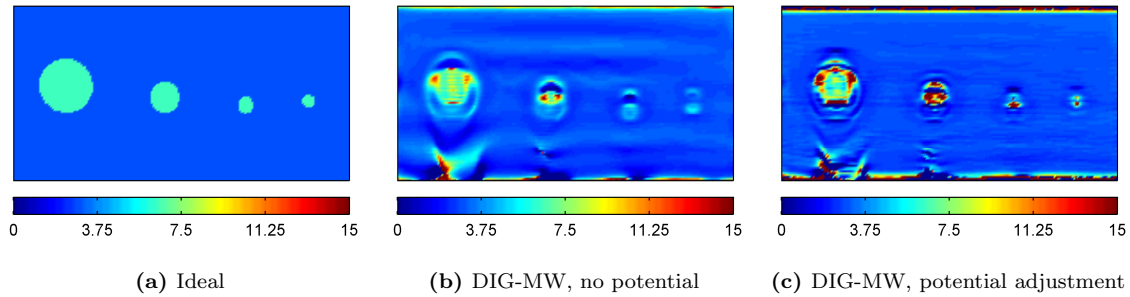
(a) Ideal   (b) DIG-MW, no potential   (c) DIG-MW, potential adjustment

Figure 5.10: A potential function, Φ, is introduced in order to allow for some variability in $\mathcal{M}$.

5.3 Summary

A new inversion method was derived in this chapter. The idea is to introduce a Green's function in order to remove one of the assumptions made in the IH inversion. The introduction of this Green's function changes the nature of the inversion from being local to global. Unfortunately, knowledge of continuous displacements outside the domain of measurement is required. To compensate for this lack of knowledge, data windows are formally introduced into the equations of motion. Due to the extra terms created by these windows, the shear modulus cannot be determined without designing an iterative procedure. For an initial estimate, it can be assumed that the shear modulus is constant over the support of the *derivatives* of the window. This is a more relaxed condition than requiring it to be constant over the entire support if the interior of the window is constant.

A set of window functions were derived that use polynomials to smoothly transition from one in the interior, to zero on the boundaries. The degree of continuity of the window determines the degree of the transition polynomials. Since these windows are to be used in discrete calculations, and since the support will be small, it is best to represent them in the spatial-frequency domain.

By introducing windows, the domain can be decomposed into a set of blocks on which the shear modulus can be estimated independently. A global solution can then be created by merging the block solutions. Noise within each block is smoothed by the convolution, and the impact of outliers and systematic errors is limited by the block size. Overlapping blocks and smoothly merging the resulting solutions can lead to a smooth global estimate.

When the window is shifted one grid point at a time, the resulting inversion process describes a convolution, which is numerically more efficient to implement than even the disjoint block inversion. With the assumption that $\mathcal{M}$ is constant over the support of the derivatives of the window, this moving-window method degenerates to Oliphant's Interior Helmholtz inversion, with $h = \mathcal{G}W$. This derivation, however, gives motivation for including the Green's function in the filter. In IH inversion, the form of filter was left somewhat arbitrary.

In an attempt to remove the local homogeneity assumption, the original equation was split into a system, where a form of potential function is introduced. This splits the task of determining the shear modulus into two steps: first solve for a potential, Φ, using the DIG method, then use this potential to find $\mathcal{M}$. The potential may allow for more variation in the shear modulus. When applied to the gel phantom data, there is a slight improvement in some regions of the estimate. However, more spikes are

introduced, and the stiffness within the cylindrical inclusions is overestimated. There may be room for improvement with this method, but it is left as future work.

It is the opinion of the author that there are still far too many assumptions and approximations made in this method. A local homogeneity assumption is made to eliminate derivatives of the shear modulus, an incompressibility assumption is made to decouple the equations of motion, and a larger homogeneity assumption is made to remove $\mathcal{M}$ from an integral expression. These are very limiting. The stiffness of a tissue is not expected to be constant. In fact, the point of the MRE procedure is to look for strong changes in stiffness that might indicate pathological tissue. Also, motions are not expected to be uncoupled. Coupling is even found in the experiment; although forcing is oriented in the z-direction, all three components of the displacement field show motion. An ideal method would not make any of these assumptions.

In the original AIDE method, only the first local homogeneity assumption is made. This is used as a starting point for the method derived in the next chapter. To allow for local variations, the derivative terms of the shear modulus are reintroduced to the system.

INVERSION BY QUADRATIC OPTIMIZATION

The largest shortcoming of the AIDE and DIG methods is the need to make homogeneity assumptions. The local homogeneity assumption is required to remove all derivative terms of $\mathcal{M}$, resulting in an algebraic expression for the modulus. It is also required to introduce the Laplacian operator and remove coupling terms, converting the system into an Interior Helmholtz equation. A larger homogeneity assumption is required to apply filters that minimize the effect of noise. Whenever there are noticeable variations in tissue stiffness, these assumptions are invalid. The entire motivation for MRE is that pathological tissues tend to have a noticeably higher stiffness than the surrounding normal tissue, which implies there are regions where the shear modulus is expected to change quite rapidly. If the homogeneity assumptions are made, the resulting estimate will be invalid in these regions. This is a major flaw in the methods. These regions of high change likely mark the boundaries of diseased tissue, so are probably the most important regions of all. The ideal method would still be fast and direct, like AIDE and DIG, but would allow for local variations in the shear modulus. The development of such a technique is the major goal in this chapter.

6.1 Quadratic Optimization

In this dissertation, *quadratic optimization* refers to problems of the following form:

$$
\begin{aligned}
&\textit{Minimize} &&f(x) = x^{\mathrm{H}}Qx + c^{\mathrm{H}}x + x^{\mathrm{H}}c,\\
&\textit{subject to} &&Ex = d.
\end{aligned}
$$

This is the Equality-constrained Quadratic Programming (EQP) problem that was discussed in Chapter 3. It is quadratic in the unknown vector x, and since one is searching for the minimizing solution, is an optimization problem. This definition includes finding least-squares solutions to an overdetermined system. Minimizing the squared error, $\|Ax - b\|^2$, is equivalent to minimizing

$$
f(x) = x^{\mathrm{H}}\left(A^{\mathrm{H}}A\right)x + \left(-A^{\mathrm{H}}b\right)^{\mathrm{H}}x + x^{\mathrm{H}}\left(-A^{\mathrm{H}}b\right),
$$

subject to the trivial constraint $(0)x = 0$. In this manner, all previously described elasticity reconstruction methods can be re-posed in this quadratic optimization framework. The advantage of this new perspective is that the minimization functional can be tweaked to more appropriately handle the assumptions. For example, instead of assuming a particular term is zero, one can search for the solution in which the magnitude of the term is minimized. This approach will be referred to as *Inversion by Quadratic Optimization* (IQO). In this way, the same assumptions can be used for regularization, but the solution is more faithful to the original inverse problem.

6.2 Pointwise Inversion by Quadratic Optimization

Before designing an inverse procedure, the system is first expanded using the product rule of differentiation. This allows one to explicitly see the terms that are ignored by AIDE:

$$-\omega^2 \mathbf{U} = \mathcal{M}\nabla \cdot \left[\nabla \mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right] + \Lambda\nabla\nabla \cdot \mathbf{U} + \underbrace{\nabla\mathcal{M} \cdot \left[\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right] + \nabla\Lambda\left(\nabla \cdot \mathbf{U}\right)}_{\text{not necessarily zero}}. \tag{6.1}$$

Equation (6.1) is equivalent to the original model, assuming a constant density. There are eight unknown terms in this equation: $\mathcal{M}$, Λ and their three first derivatives. By expressing the equation in this expanded form, more can be said about the effect of the 'simplifications' described in Section 3.1.8. If the longitudinal component is ignored, then Λ and $\nabla\Lambda$ are zero, leaving only four unknowns: $\mathcal{M}$, $\frac{\partial\mathcal{M}}{\partial x}, \frac{\partial\mathcal{M}}{\partial y}$ and $\frac{\partial\mathcal{M}}{\partial z}$. If the longitudinal component is replaced with a pressure term, then the equation becomes

$$-\omega^2 \mathbf{U} = \mathcal{M}\nabla \cdot \left[\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right] + \nabla\mathcal{M} \cdot \left[\nabla\mathbf{U} + \nabla^{\mathrm{T}}\mathbf{U}\right] + \nabla P,$$

which only has seven unknown terms: $\mathcal{M}$, $\frac{\partial\mathcal{M}}{\partial x}, \frac{\partial\mathcal{M}}{\partial y}, \frac{\partial\mathcal{M}}{\partial z}, \frac{\partial P}{\partial x}, \frac{\partial P}{\partial y}$ and $\frac{\partial P}{\partial z}$. If the curl operator is applied to the equation of motion, the longitudinal components are removed, but all second derivatives of $\mathcal{M}$ are introduced, increasing the net number of unknown terms to ten. This suggests that applying the curl may not be an appropriate course of action in order to remove the undesired parameter.

To keep the problem relatively small, it is assumed that the longitudinal wave component is negligible, but no assumptions are made about the shear modulus. The resulting system, in matrix form, is the following:

$$\begin{bmatrix} U_{1,ii} + U_{i,i1} & 2U_{1,1} & U_{1,2} + U_{2,1} & U_{1,3} + U_{3,1} \\ U_{2,ii} + U_{i,i2} & U_{2,1} + U_{1,2} & 2U_{2,2} & U_{2,3} + U_{3,2} \\ U_{3,ii} + U_{i,i3} & U_{3,1} + U_{1,3} & U_{3,2} + U_{2,3} & 2U_{3,3} \end{bmatrix} \begin{bmatrix} \mathcal{M}_0 \\ \mathcal{M}_1 \\ \mathcal{M}_2 \\ \mathcal{M}_3 \end{bmatrix} = -\omega^2 \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}, \tag{6.2}$$

where $\mathcal{M}_0$ is an estimate of $\mathcal{M}$, and $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ is an estimate of $\nabla\mathcal{M}$. In practice, length scales should be introduced so that each $\mathcal{M}_i$ is likely to be of the same order. This can be accomplished by setting

$$\tilde{\mathcal{M}}_0 = \mathcal{M}_0, \quad \tilde{\mathcal{M}}_1 = \Delta x\, \mathcal{M}_1, \quad \tilde{\mathcal{M}}_2 = \Delta y\, \mathcal{M}_2, \quad \tilde{\mathcal{M}}_3 = \Delta z\, \mathcal{M}_3.$$

From this point on, it is assumed that this rescaling is performed.

At every point, Equation (6.2) is a set of three equations for four unknown terms. This suggests that if four linearly independent equations are generated, then all four terms can be estimated as independent functions. Note that the coefficients of the unknowns are related to the displacement field. If a second linearly independent displacement field is known for the same tissue, more equations can be generated. A second field can be obtained by repeating the experiment, changing the orientation or location of the external stress. The more times the experiment is repeated, the more equations generated. This allows for a unique determination of $\mathcal{M}$, even though Equation (6.2) is a system of differential equations. No boundary conditions are required. By creating large systems of differential equations for few unknown

functions, unique solutions can be obtained without any form of external conditions. This concept is described further in Appendix C. The expanded system allows for local variations in $\mathcal{M}$, removing the homogeneity assumption. Furthermore, since the equations are still entirely local, a pointwise inversion can be performed. This method will be referred to as *Pointwise Inversion by Quadratic Optimization* (PIQO).

This pointwise approach treats $\{\mathcal{M}_i\}$ as independent functions. If all four are independently estimated, then a smoothed version of $\mathcal{M}$ can be reconstructed by finding the function that minimizes

$$f(\mathcal{M}) = \|\mathcal{M} - \mathcal{M}_0\|^2 + \alpha_1 \|D_1\mathcal{M} - \mathcal{M}_1\|^2 + \alpha_2 \|D_2\mathcal{M} - \mathcal{M}_2\|^2 + \alpha_3 \|D_3\mathcal{M} - \mathcal{M}_3\|^2, \qquad (6.3)$$

where $\{D_i\}$ are discrete derivative operators, and $\{\alpha_i\}$ are weights. If further regularization is desired, a regularization term can be added to this functional, perhaps limiting the variability of $\mathcal{M}$.

For the PIQO method to be applied, there must be at least four linearly independent equations. By performing the experiment only once, a maximum of three equations are generated. Thus, the experiment should be repeated at least once. Unfortunately, this concept could not be tested with the gel phantom data used in the previous chapters; the resources to perform the experiment were unavailable, and experimental data was only provided for one instance of the experiment. For this reason, displacement fields were simulated using a technique described in Appendix B. The tissue was chosen to be perfectly elastic, and the longitudinal component of the wave was ignored. A spherical inclusion was placed in the centre of the tissue, with a relatively smooth transition to prevent reflections. One side of the tissue was excited with sinusoidal motion oriented in the z-direction. A slice of this field is shown in Figure 6.1.
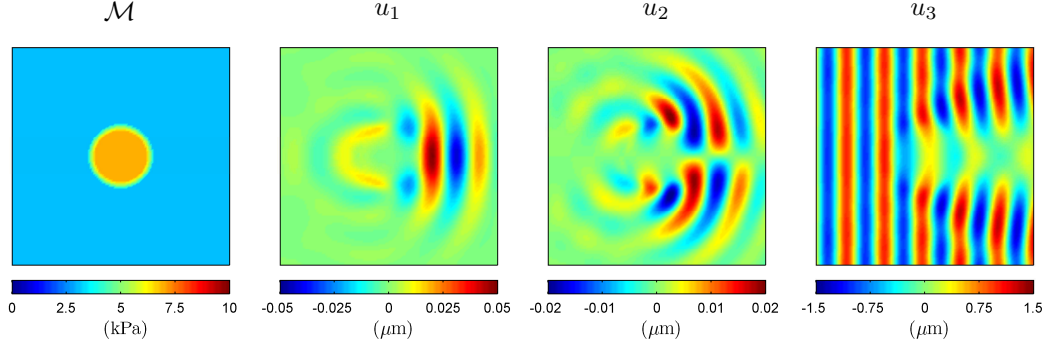

Figure 6.1: Simulated displacement field for a spherical inclusion. Forcing is applied to the left side of the tissue, oriented in the z-direction. Images show the centre slice, parallel to the xy-plane.

Figure 6.2 shows estimates of the shear modulus and its derivatives using displacement data from two and four simulations. The PIQO method accurately estimates $\mathcal{M}$ and all its first derivatives, even though all four parameters are estimated as independent functions. The more data available, the more accurate the estimates. There are some discrepancies due to numerical noise, but these can be removed in post-processing by using the functional in Equation (6.3) to find a smooth version of the shear modulus.

If no extra experiments are performed, the pointwise system remains underdetermined, and cannot be solved without further assumptions. No intelligent assumptions can be made relating the values of
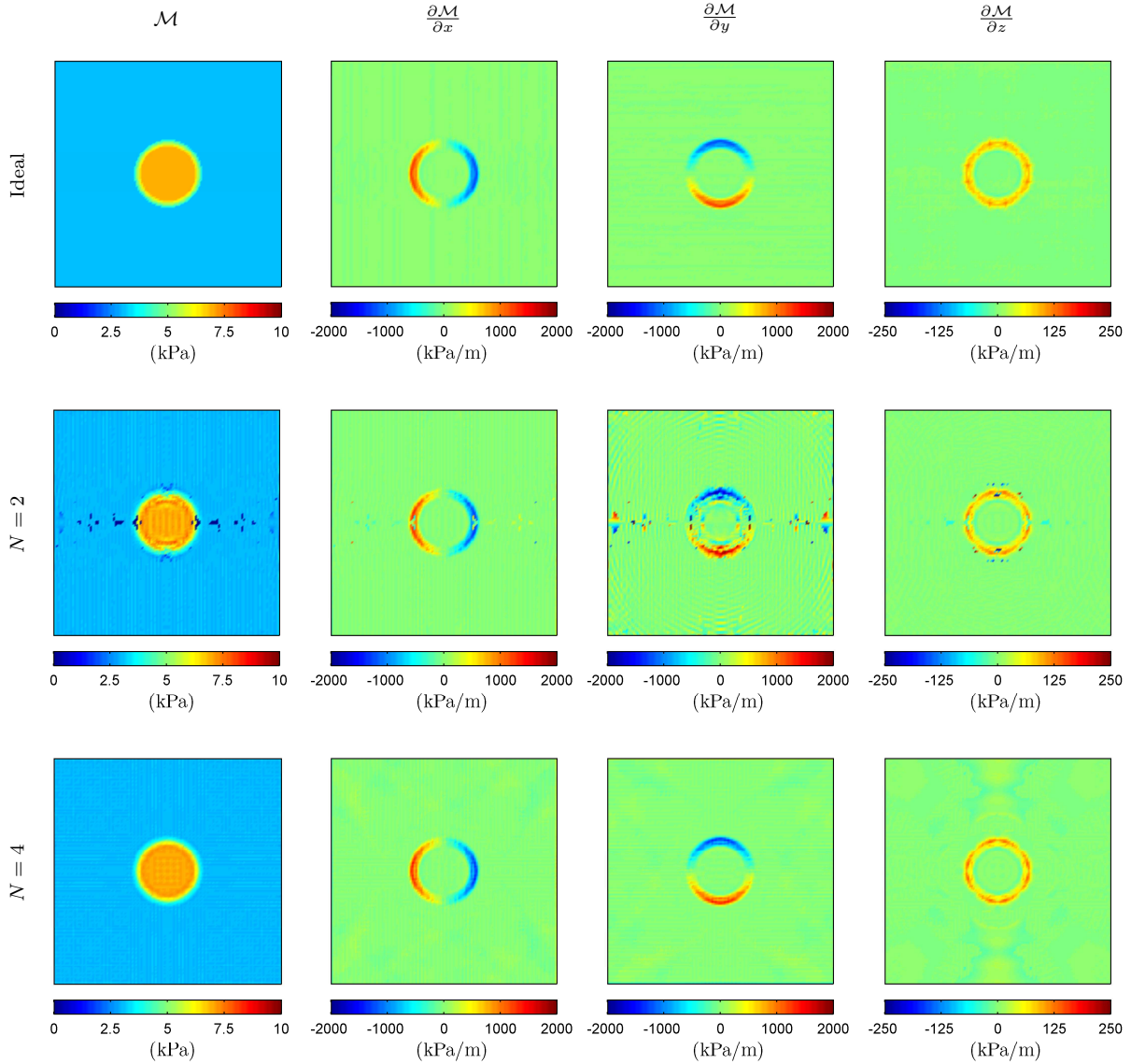
Figure 6.2: PIQO method with simulated data from two and four linearly independent displacement fields. For the two-field case, forcing is applied on the left and right ($x = x_l$, $x = x_r$). For the four-field inversion, forcing is applied on the four visible sides ($x = x_l$, $x = x_r$, $y = y_l$, $y = y_r$).

(a) $\mathcal{M}_0$ **(b)** $\mathcal{M}_1$ **(c)** $\mathcal{M}_2$ **(d)** $\mathcal{M}_3$
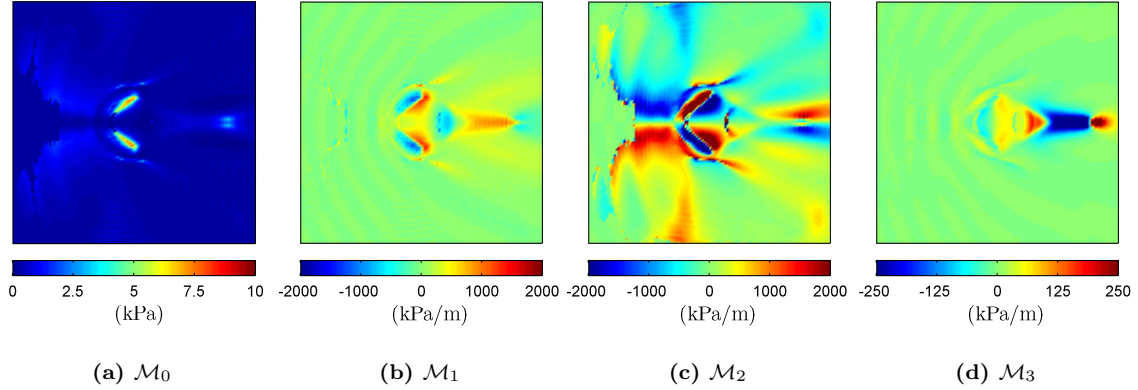
Figure 6.3: PIQO estimate from a single displacement field. The norm of $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ was minimized in order to steer the system to a unique solution. Forcing was applied to the left side of the simulated tissue.

$\mathcal{M}_0(\boldsymbol{x}_o)$, $\mathcal{M}_1(\boldsymbol{x}_o)$, $\mathcal{M}_2(\boldsymbol{x}_o)$ and $\mathcal{M}_3(\boldsymbol{x}_o)$, without including information from surrounding points. For example, one cannot attempt to set $\mathcal{M}_1 = D_1 \mathcal{M}_0$ without knowing encompassing values of $\mathcal{M}_0$. It is not logical to minimize any of the four functions without including the derivative relationship between them. These relations can be used, however, if the inversion is performed on blocks of data.

As a demonstration of the weakness of PIQO for an underdetermined system, the method was applied to the data from a single field, and the resulting shear modulus estimate is shown in Figure 6.3. For this method, the magnitude of $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ was minimized. The estimate is poor because there are too many free parameters in the system. This emphasizes the importance of having additional information.

6.3 Blockwise Inversion by Quadratic Optimization

In order to relate the unknown functions, $\{\mathcal{M}_i\}$, a collection of points must be considered at once. This is the idea behind *Blockwise Inversion by Quadratic Optimization* (BIQO). Considering a block of points allows one to discretely approximate the derivatives of $\mathcal{M}_0$. In matrix form, the relationship between the functions is

$$\begin{bmatrix} -D_1 & I & 0 & 0 \\ -D_2 & 0 & I & 0 \\ -D_3 & 0 & 0 & I \end{bmatrix} \begin{bmatrix} \mathcal{M}_0 \\ \mathcal{M}_1 \\ \mathcal{M}_2 \\ \mathcal{M}_3 \end{bmatrix} = 0, \tag{6.4}$$

where $\mathcal{M}_0$ is now a vector representation of the shear modulus for a rasterized block of data. This system enforces the derivative relationships. The equations of motion are also converted into blocks:

$$
\begin{bmatrix}
\hat{U}_{1,ii} + \hat{U}_{i,i1} & 2\hat{U}_{1,1} & \hat{U}_{1,2} + \hat{U}_{2,1} & \hat{U}_{1,3} + \hat{U}_{3,1} \\
\hat{U}_{2,ii} + \hat{U}_{i,i2} & \hat{U}_{2,1} + \hat{U}_{1,2} & 2\hat{U}_{2,2} & \hat{U}_{2,3} + \hat{U}_{3,2} \\
\hat{U}_{3,ii} + \hat{U}_{i,i3} & \hat{U}_{3,1} + \hat{U}_{1,3} & \hat{U}_{3,2} + \hat{U}_{2,3} & 2\hat{U}_{3,3}
\end{bmatrix}
\begin{bmatrix}
\mathcal{M}_0 \\
\mathcal{M}_1 \\
\mathcal{M}_2 \\
\mathcal{M}_3
\end{bmatrix}
= -\omega^2
\begin{bmatrix}
U_1 \\
U_2 \\
U_3
\end{bmatrix},
\tag{6.5}
$$

where U_k is a rasterized vector of displacements, and $\hat{U}_k = \mathrm{diag}\{U_k\}$ is a square diagonal matrix. Equation (6.5) implements the pointwise equations in a block matrix form, but now allows the derivative relationships to be applied. This can be accomplished in two ways: combining Equations (6.4) and (6.5) into a single equality constraint, or implementing the derivative relationships in a quadratic functional to be minimized.

If the shear modulus and its derivatives are estimated on disjoint blocks, the boundaries of these subdomains tend to appear in the solution. To prevent this, overlapping blocks and smooth merging can be used, exactly as was done for the DIG method in Section 5.1.4.

There are instances when some of the the row equations in (6.5) are effectively linearly dependent. This can make the system ill-conditioned. Thus, before proceeding, the system should be reduced to an effective system. This was accomplished using the singular value decomposition technique described in Section 3.3.3.2. Once a well-conditioned, reduced system is determined, the quadratic programming techniques can be applied.

6.3.1 Equality Constrained (BIQO–E)

When the derivative relationships are added to the set of equality constraints, they must be weighted in order to properly balance the contributions of Equations (6.4) and (6.5):

$$
\begin{bmatrix}
\hat{U}_{1,ii} + \hat{U}_{i,i1} & 2\hat{U}_{1,1} & \hat{U}_{1,2} + \hat{U}_{2,1} & \hat{U}_{1,3} + \hat{U}_{3,1} \\
\hat{U}_{2,ii} + \hat{U}_{i,i2} & \hat{U}_{2,1} + \hat{U}_{1,2} & 2\hat{U}_{2,2} & \hat{U}_{2,3} + \hat{U}_{3,2} \\
\hat{U}_{3,ii} + \hat{U}_{i,i3} & \hat{U}_{3,1} + \hat{U}_{1,3} & \hat{U}_{3,2} + \hat{U}_{2,3} & 2\hat{U}_{3,3} \\
-\alpha D_1 & \alpha I & 0 & 0 \\
-\alpha D_2 & 0 & \alpha I & 0 \\
-\alpha D_3 & 0 & 0 & \alpha I
\end{bmatrix}
\begin{bmatrix}
\mathcal{M}_0 \\
\mathcal{M}_1 \\
\mathcal{M}_2 \\
\mathcal{M}_3
\end{bmatrix}
= -\omega^2
\begin{bmatrix}
U_1 \\
U_2 \\
U_3 \\
0 \\
0 \\
0
\end{bmatrix}.
\tag{6.6}
$$

This potentially creates an overdetermined system, which can be solved using least-squares. In least-squares, the relative magnitudes of the coefficients determine which equations are favoured, which is why the weighting factor, α, is important. If α is too large, the derivative relations will hold, but the measured data will be poorly fit. If α is too small, the data will be well-fit, but the derivative relationships will be poorly satisfied. Depending on the number of free parameters, the system can be sensitive to changes in α. The optimal value depends on the confidence in the data, and the confidence in the discrete approximations of the derivatives.

The new system might still be effectively underdetermined, which means further regularization might be required. Since the derivative relationships are enforced, it is now appropriate to minimize the

gradient, $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$, by using

$$
Q = \begin{bmatrix} \beta I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}
$$

in the quadratic minimization procedure. The weight, β, is a small parameter that increases stability of the algorithm. In this blockwise approach, the $\mathcal{M}_i$'s are no longer independent, which means minimizing the norms of $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ *does* affect the smoothness of $\mathcal{M}_0$. This was not the case in the pointwise approach. Another appropriate choice might be to minimize all gradients:

$$
Q = \begin{bmatrix} D & 0 & 0 & 0 \\ 0 & D & 0 & 0 \\ 0 & 0 & D & 0 \\ 0 & 0 & 0 & D \end{bmatrix}, \quad \text{where } D = D_1^{\text{H}} D_1 + D_2^{\text{H}} D_2 + D_3^{\text{H}} D_3.
$$

This will attempt to minimize the variation in all functions, resulting in smooth solutions.

In practice, this further regularization was not required if Equation (6.5) contained at least two effectively linearly independent equations at every point. This suggests that two differential equations at each point are enough to have a fully determined system, even though no boundary conditions are applied. Possible explanations for this are explored in Appendix C. For large block sizes, further regularization using standard null-space and range-space methods is a very slow process. A possible future direction is to develop a fast, iterative methods of performing the added regularization. For now, since it was found that further regularization was not usually required, least-squares alone is used to solve Equation (6.6).
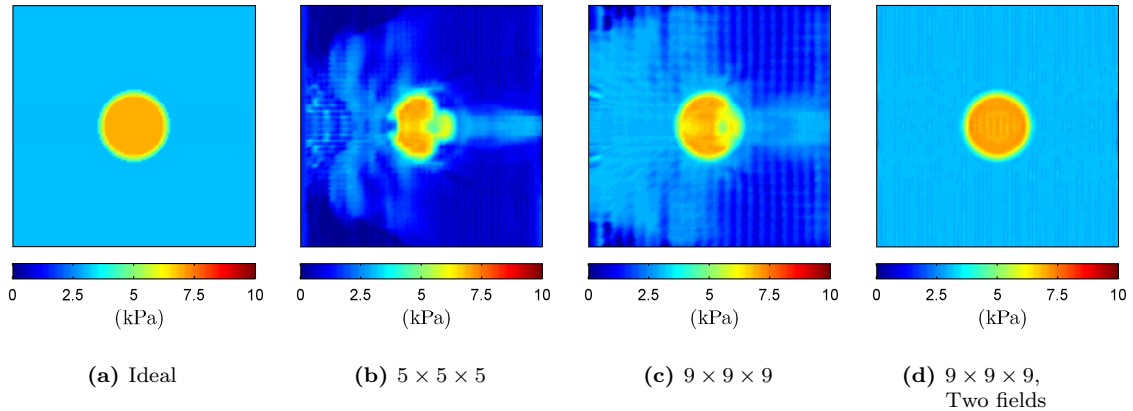

Figure 6.4: BIQO–E method with simulated data. The tuning parameter was chosen to have the value $\alpha = 0.01$. The first two estimates use data from one simulation, where forcing was applied on the left. The third estimate uses data from two simulations, where forcing was applied separately on the left and right.

This method, referred to as BIQO–E, was applied to the simulated data. Results are displayed in Figure 6.4. The solution is shown to depend on block size. Larger blocks have more accurate approximations in the derivative relations, resulting in a more accurate solution. If data from two fields are used, the shear modulus estimate is greatly improved.

There are a few disadvantages of using the derivative relations as part of the equality constraint. First of all, the system of equations is much larger than in the pointwise method, which puts more of a strain on computing resources. For reasonable computation times, the blocks of data should be relatively small, which limits the potential accuracy of the discrete derivative approximations. The sensitivity to the α parameter is also undesirable. Just as for Tikhonov regularization methods, different choices of α may produce different estimates.

6.3.2 Quadratically Minimized (BIQO–Q)

Instead of adding equations to the equality constraint, the derivative relations between $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ can be used in the minimization functional. The parameters in the EQP problem for this case are the following:

$$E = \begin{bmatrix} \hat{U}_{1,ii} + \hat{U}_{i,i1} & 2\hat{U}_{1,1} & \hat{U}_{1,2} + \hat{U}_{2,1} & \hat{U}_{1,3} + \hat{U}_{3,1} \\ \hat{U}_{2,ii} + \hat{U}_{i,i2} & \hat{U}_{2,1} + \hat{U}_{1,2} & 2\hat{U}_{2,2} & \hat{U}_{2,3} + \hat{U}_{3,2} \\ \hat{U}_{3,ii} + \hat{U}_{i,i3} & \hat{U}_{3,1} + \hat{U}_{1,3} & \hat{U}_{3,2} + \hat{U}_{2,3} & 2\hat{U}_{3,3} \end{bmatrix}, \qquad d = -\omega^2 \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix},$$

$$Q = \begin{bmatrix} D & -D_1^{\mathrm{H}} & -D_2^{\mathrm{H}} & -D_3^{\mathrm{H}} \\ -D_1 & I & 0 & 0 \\ -D_2 & 0 & I & 0 \\ -Dz & 0 & 0 & I \end{bmatrix}, \qquad c = 0.$$

This forces the estimates of the shear modulus and its derivatives to match the data, first, then attempts to minimize

$$f\left(\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\right) = \|D_1\mathcal{M}_0 - \mathcal{M}_1\|^2 + \|D_2\mathcal{M}_0 - \mathcal{M}_2\|^2 + \|D_3\mathcal{M}_0 - \mathcal{M}_3\|^2$$

to determine any free parameters. If further regularization is required, Q can be modified to add another minimization term. Again, this regularization is very slow for large block sizes. Faster optimization methods should be explored.

The BIQO–Q method was applied to the simulated data. Results are shown in Figure 6.5. Just as for BIQO–E, the solution depends on the block size because of the discrete derivative approximations. Larger blocks have more accurate derivative relations, resulting in a more accurate solution. If data from two fields are used, the resulting estimate is greatly improved. The BIQO–Q method does not involve the potentially sensitive parameter, α. The solution is always guaranteed to closely match the data. This might be an issue if the data contains noise or does not closely fit the assumptions, such as ignoring the longitudinal component. Comparing Figure 6.5 to Figure 6.4, it seems that including the derivative relations in the equality constraint results in a better estimate for a fixed block size. BIQO–E is also faster if the system is determined, since least-squares can be used. Thus, BIQO–Q is concluded to be inferior.
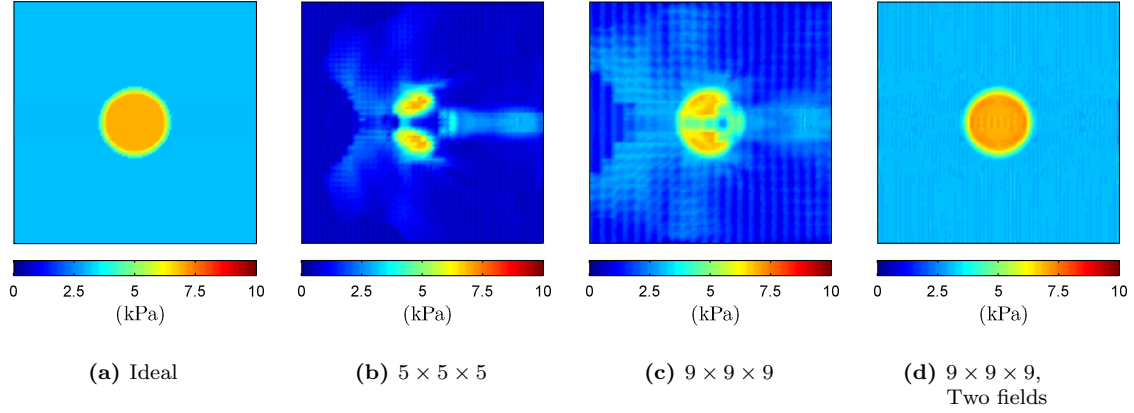
(a) Ideal   (b) $5 \times 5 \times 5$ (c) $9 \times 9 \times 9$ (d) $9 \times 9 \times 9$, Two fields

Figure 6.5: BIQO–Q method with simulated data. The first two estimates use data from one simulation, where forcing was applied on the left. The third estimate uses data from two simulations, where forcing was applied separately on the left and right.

6.3.3   Single Parameter (BIQO–S)

If the discrete derivative relations are expected to hold perfectly, then $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ can be directly replaced with $(D_1\mathcal{M}_0, D_2\mathcal{M}_0, D_3\mathcal{M}_0)$, resulting in an expression with a single unknown function:

$$\begin{bmatrix} \hat{U}_{1,ii} + \hat{U}_{i,i1} & 2\hat{U}_{1,1} & \hat{U}_{1,2} + \hat{U}_{2,1} & \hat{U}_{1,3} + \hat{U}_{3,1} \\ \hat{U}_{2,ii} + \hat{U}_{i,i2} & \hat{U}_{2,1} + \hat{U}_{1,2} & 2\hat{U}_{2,2} & \hat{U}_{2,3} + \hat{U}_{3,2} \\ \hat{U}_{3,ii} + \hat{U}_{i,i3} & \hat{U}_{3,1} + \hat{U}_{1,3} & \hat{U}_{3,2} + \hat{U}_{2,3} & 2\hat{U}_{3,3} \end{bmatrix} \begin{bmatrix} I \\ D_1 \\ D_2 \\ D_3 \end{bmatrix} \mathcal{M}_0 = -\omega^2 \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}.$$

This has the advantage of being a much smaller system, so larger block sizes can be used. The disadvantage is that there is less control. In BIQO–E, α can tune the effect of the derivative constraint, based on confidence in the data and the discrete derivative approximation. If the block sizes are small, poor discrete derivative estimates can make this method very sensitive to noise.

This inversion method was applied to the simulated data. The resulting shear modulus estimates are shown in Figure 6.6. Again, block size directly affects the solution. This method is much faster than the previous ones, but results are not as accurate as those obtained by BIQO–E for a given block size. This method also does not directly estimate the derivatives of the shear modulus, which is done by the other two methods. These derivatives can give strong indications of where inclusions are found. For BIQO–S, the derivatives must be estimated from $\mathcal{M}_0$, after the inversion.

6.4   Decoupled Inversion (BIQO–D)

The methods described in this chapter, thus far, require all three polarizations of motion for each experiment. As described in Section 3.2.4, each polarization is captured in a separate image during data acquisition. This can lead to long imaging times, particularly if the experiment is repeated for several
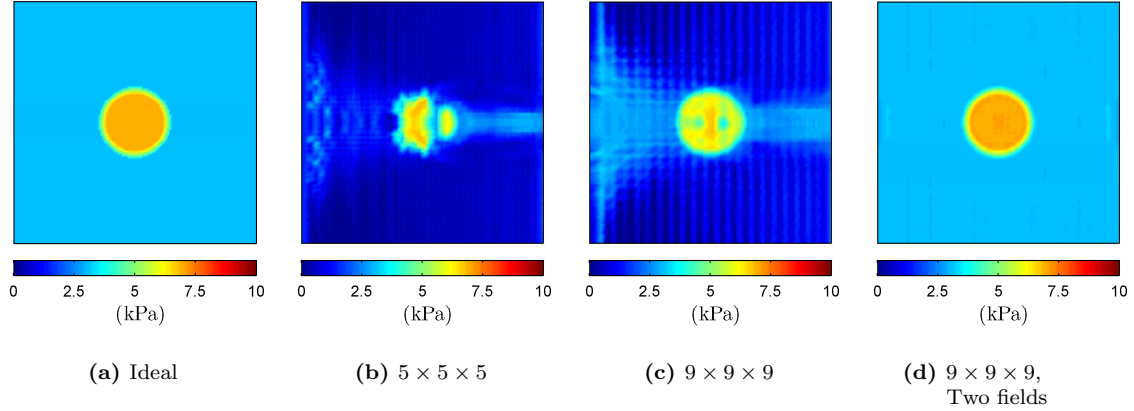
(a) Ideal   (b) $5 \times 5 \times 5$ (c) $9 \times 9 \times 9$ (d) $9 \times 9 \times 9$,
                                                                                   Two fields

Figure 6.6: BIQO–S method with simulated data. The first two estimates use data from one simulation, where forcing was applied on the left. The third estimate uses data from two simulations, where forcing was applied separately on the left and right.

locations of external forcing. If a reasonable estimate could be obtained using a single polarization, the total imaging time can be drastically reduced. Repeating the experiment then becomes much more cost-effective.

By ignoring the coupling terms, the system with repeated experiments is the following:

$$
\begin{bmatrix} \hat{U}_{j,ii} & \hat{U}_{j,1} & \hat{U}_{j,2} & \hat{U}_{j,3} \end{bmatrix}
\begin{bmatrix} \mathcal{M}_0 \\ \mathcal{M}_1 \\ \mathcal{M}_2 \\ \mathcal{M}_3 \end{bmatrix} = -\omega^2 \begin{bmatrix} U_j \end{bmatrix} ,
$$

where U_j is the rasterized vector of displacements from the jth experiment, oriented in the direction of the forcing. With this equation of motion, the concepts behind BIQO can be applied. In what follows, the derivative relations are included as part of the equality constraint, and the system is solved using least squares if there are at least two polarizations available.

The BIQO–D method was applied to the simulated data, with forcing from one, two, and four different locations. The resulting shear modulus estimates are shown in Figure 6.7. From this figure, it is apparent that a relatively smooth estimate can be obtained from only two experiments, with a single polarization of motion from each. Adding information from further fields only aids in smoothing the estimate. The value within the spherical inclusion seems to be underestimated, however. This results from a loss of the coupling terms in the equations, so cannot be improved by further experiments. This may not be an issue in practice. If one only wishes to find regions of higher stiffness, then the precise value of the shear modulus may not be important. This removal of the coupling terms might, however, have a stronger impact with more complicated geometries. This should be tested in practice.
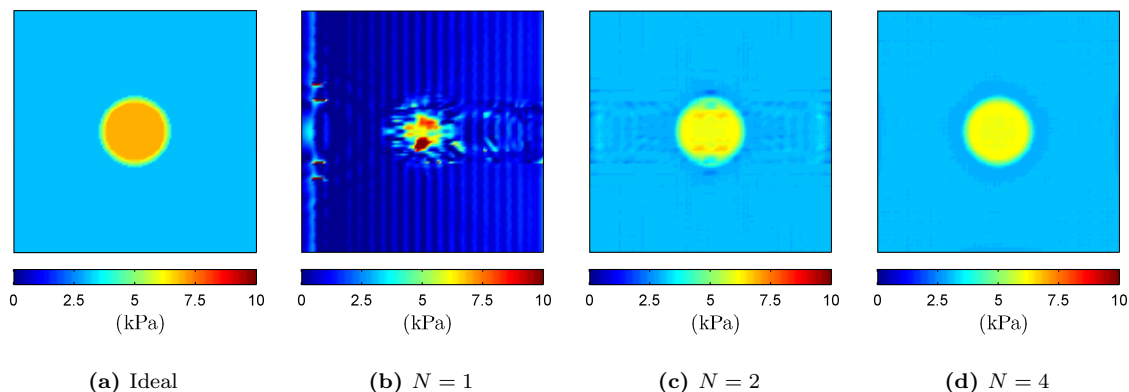
114

(a) Ideal   **(b)** $N = 1$ **(c)** $N = 2$ **(d)** $N = 4$

Figure 6.7: BIQO–D estimate with N polarizations of motion. The block size is $9 \times 9 \times 9$. Only the polarization of motion in the orientation of the forcing is used, and estimates are obtained by combining motions from one, two and four separate experiments with differing locations of forcing.

6.5   Summary

The goal of this chapter was to develop inversion methods that do not rely on the commonly made local homogeneity assumption. Instead of regularizing the system by changing the equations of motion, the problem is re-posed in a quadratic optimization framework, and regularization is performed by adding terms to a steering functional. This is the idea behind Inversion by Quadratic Optimization (IQO).

The equations were expanded and rewritten in matrix form, creating a system with four unknown functions at every point: $\mathcal{M}$ and its three first derivatives. It was observed that with additional equations, which can be obtained by repeating the experiment, all four functions can be estimated, without considering the derivative relationships between them. This is a pointwise method, so was labelled Pointwise Inversion by Quadratic Optimization (PIQO). With fewer than four equations, some form of regularization is required. However, without forcing the relationship between $\mathcal{M}$ and its derivatives, no logical regularization can be performed. To include the derivative relationships, the shear modulus must be inverted in blocks.

Three Blockwise Inversion by Quadratic Optimization (BIQO) methods were presented that use the coupled equations of motion. In BIQO–E, the derivative relations are added to the system of equality constraints. In BIQO–Q, these relations are used in the quadratic functional. In BIQO–S, the derivative relations are directly substituted into the equations of motion. Each method has its own set of advantages and disadvantages. However, BIQO–E is essentially a generalized version of the other two. With $\alpha \gg 1$, BIQO–E $\approx$ BIQO–S, and with $\alpha \ll 1$, BIQO–E $\approx$ BIQO–Q. The tuning parameter, α, can be used to balance the confidence in the data and the confidence in the discrete approximations of the derivative operator. This leads to better results for a given block size. Results were shown to improve as the block size increases. This is explored further in the next chapter. Results were also found to improve when data was used from multiple experiments, with linearly independent displacement fields.

Capturing all three polarizations of motion is costly; in the data acquisition procedure, each polar-

ization is obtained in a separate image. To generate each image, the entire volume must be scanned more than once. In order to reduce the total scan time, the decoupling assumption is made, allowing an estimate to be obtained from a single polarization. The BIQO method was adapted to this decoupled system (BIQO–D). With only two polarizations, one each from two experiments, a smooth estimate was generated. The accuracy, however, is dependent on the validity of the decoupling assumption. For the simulated spherical inclusion data, the decoupling assumption caused the method to underestimate the stiffness value in the center of the inclusion.

These IQO methods are fundamentally different than the other methods in this dissertation. In the AIDE, optimal AIDE, and DIG methods, regularization is performed before the estimation procedure. In IQO, regularization is performed *during* inversion, and *only* if it is required. If the system of differential equations is fully determined, which seems to be the case when at least two equations are available, then no regularization is necessary, allowing for a very accurate estimate of the true shear modulus. This is not possible with the other methods.

RESULTS AND ANALYSIS

Several methods for estimating the shear modulus have been described in the three preceding chapters. These can be separated into three classes: *Algebraic Inversion of the Differential Equation* (AIDE), *Direct Inversion by Green's functions* (DIG), and *Inversion by Quadratic Optimization* (IQO). In each chapter, the algorithms were tested with an example displacement field to highlight key features of the inversions. Here, the methods are compared to each other. Four techniques in particular are examined: AIDE, AOSMD, DIG_1–MW, and BIQO–E. These are tested for both accuracy and robustness to the presence of noise.

Performance can often be improved by first 'cleaning' the data through filtering or smoothing, removing as much noise or other undesired information as possible. This is known as pre-processing. Many such techniques were described in Chapter 3. Of these, the Savitsky-Golay smoothing filter, the Butterworth bandpass filter, and the directional filter are tested. The first two are mainly used for eliminating noise, and the last is used to remove waves caused by reflections. The directional filter changes the nature of the data, making it non-physical. The full implications of this are explored.

The last comparisons involve repeated experiments. In the development of the IQO methods, it is suggested that accuracy and stability can be increased if data from a second, linearly independent displacement field is used. A second field can be generated by changing the location of the external forcing. This idea is tested for all four methods.

The comparisons are performed on three test-cases. Two of these have already been seen: the gel phantom displacement field, and the simulated spherical inclusion field. A third test-case is simulated, in which the shear modulus is spatially varying over the entire domain (i.e. not constructed by piecewise constants). These three data sets are described, in detail, in the next section.

7.1 The Displacement Data

The input to the MRE inversion algorithms is a set of discrete measurements from a three-dimensional harmonic displacement field. In order to compare the accuracy of the methods, the true distribution of the shear modulus must be known. To compare the performance in the presence of noise, either noise must already exist in the data, or must be added using an appropriate noise model.

There are three displacement fields considered in this dissertation. Two of these are generated by numerical simulation, and one is obtained by experiment. The simulated data is considered relatively noise-free, so in order to test noise performance, this data is corrupted with additive white Gaussian noise. The data from experiment is considered to already contain noise, so to test accuracy of the methods, the data is pre-processed.

7.1.1   Simulated Data

In order to generate simulated data for use in MRE, the forward problem must be solved for a given map of the shear modulus. Because of the three-dimensional nature of the problem, and the large spatial grid, solving the implicit system of harmonic equations directly is very computationally expensive and challenging. Instead, a fully-explicit method was developed to solve the time-dependent equations:

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \nabla \cdot \left(\mu \left[\nabla \boldsymbol{u} + \nabla^{\mathrm{T}} \boldsymbol{u} \right] \right) + \nabla \left(\lambda \nabla \cdot \boldsymbol{u} \right), \tag{7.1}$$

where $\boldsymbol{u}$ is the time-dependent displacement field, ρ is the density, μ is the purely elastic shear modulus, and λ is the real version of Lamé's first parameter. The viscous component was removed to maintain the explicit nature of the scheme. To guarantee that a periodic quasi-steady state is reached, perfectly absorbing boundary conditions were implemented. The system was simulated until all transient behaviour had left the domain, at which point the displacement values were recorded over one full period. The method is described in detail in Appendix B.

To remove the impact of longitudinal waves, λ was taken to be zero. The density was assumed constant at 1 g/cm^3. Simulations were performed for two maps of elasticity. For the first simulation, a stiff spherical inclusion was introduced in the centre of the domain. The shear modulus is mostly constant, except in the transition region between the background and the stiff inclusion. This is the data set that was used in Chapter 6. For the second simulation, sinusoids were superimposed on a background stiffness, and a Gaussian bump was introduced in the centre of the domain. This simulated tissue was designed to test the effect of the local homogeneity assumption when the shear modulus is constantly varying. Sinusoidal forcing with a frequency of 100 Hz was applied to one of the sides of the tissue, oriented in the z-direction. The spatial grid was selected to be $128 \times 128 \times 24$ points large. The resolution was chosen so that the total volume size of the simulated tissue is $10 \times 10 \times 4.8$ cm, which is on the same order as that for the gel-phantom experiment. The field was saved at ten evenly spaced points in time over a full period.

A slice of the shear modulus and displacement fields for the two simulations is shown in Figure 7.1. Note the difference in scale of the three components of motion. The displacements oriented in the direction of the forcing have magnitudes at least ten times larger than those in the other two orthogonal polarizations. Also note the direction of wave propagation. The u_3 polarizations contain waves mostly travelling from left to right. The other two have components travelling in other directions. This becomes important when considering directional filtering. Lastly, notice the wakes behind the stiff regions. These regions of low displacement have very small signal-to-noise ratios. To combat this, the tissue should be excited from multiple sides. This can be accomplished by applying the force to multiple locations at once, or by repeating the experiment, changing the location of the external force.

The resulting simulated displacement fields contain numerical noise from the discrete approximations. Some of this can be removed by applying a smoothing filter. The noise-free displacement field is taken to be a smoothed version of the simulated field. For the smoothing, a third-order $7 \times 7 \times 7$ Savitsky-Golay smoothing filter was used.
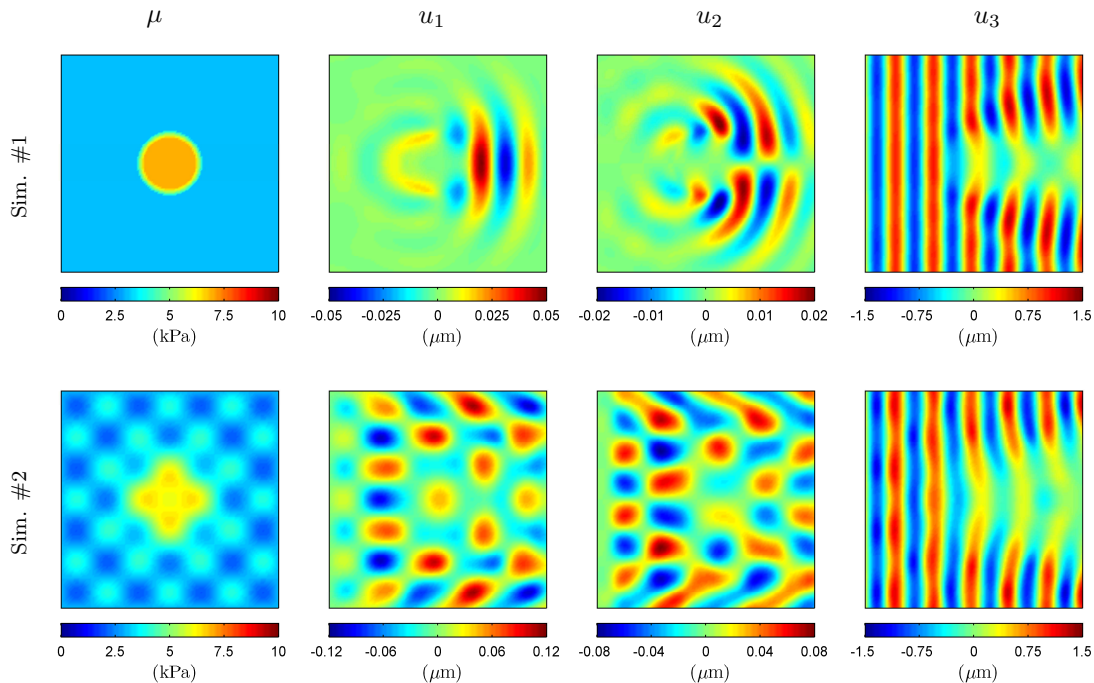
Figure 7.1: Fields are generated for two different shear modulus distributions. Forcing is applied to the left side of the tissue in the z-direction (out of page). Images show the centre slice of the shear modulus, $\mathcal{M}$, and the three resulting polarizations of motion: u_1, u_2, and u_3.

7.1.2 Measured Data

The resources were not available to perform the MR data acquisition procedure. Thankfully, a measured data set was provided by a colleague. This data set was captured by the Mayo group in Dr. Richard Ehman's lab [13], and was also used by Manduca *et al.* [39] and Oliphant [49]. The experiment was performed on a gel phantom containing four stiff cylindrical inclusions. The background gel was composed of 1.5% agar, and has a shear modulus of approximately 2.9 kPa. The inclusions were composed of 10% B-gel (bovine), and have a shear modulus of approximately 6.4 kPa. The sizes of the inclusions are 5, 10, 16 and 25 mm in diameter. Data was acquired by multi-slice 2D imaging, with a 20×20 cm field of view, and a slice thickness of 3 cm. The total spatial grid has dimensions $256 \times 256 \times 16$ points, and displacements were captured at 8 time-offsets. Forcing was oriented in the z-direction with a frequency of 100 Hz.

Since the phantom was smaller than the field of view, and the boundaries of the displacement field contained sharp peaks and outliers, the data was trimmed to a $188 \times 98 \times 14$ grid. A slice of the shear modulus and displacement field is shown in Figure 7.2. This data contains measurement noise, so must be filtered in order to estimate a noise-free field. Following Manduca *et al.* [39], the data is filtered with a directional filter to remove reflected waves. This filter was designed to admit waves travelling in the
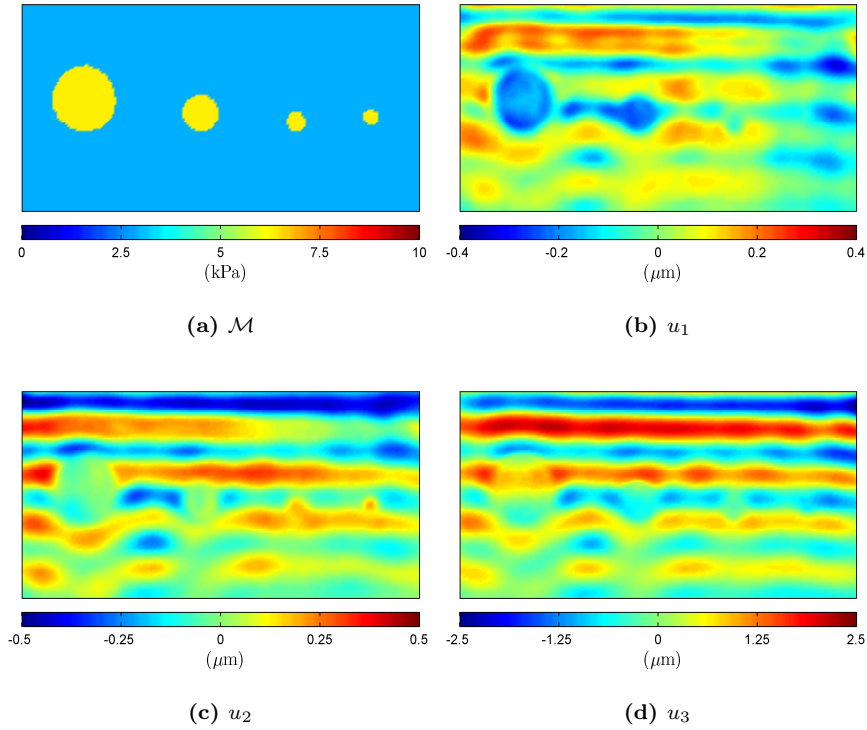
(a) $\mathcal{M}$

(b) u_1

(c) u_2

(d) u_3

Figure 7.2: Measured displacement field from the gel phantom experiment. The tissue is excited in the z-direction at the top of the figure. Images show the centre slice of the shear modulus, $\mathcal{M}$, and the displacement field components: u_1, u_1, and u_3.

direction of

$$\hat{\boldsymbol{n}} = (-0.0161, 0.9999, 0.0051).$$

This vector is the best estimate of the direction in which the majority of the waves are propagating. It was computed by fitting a line through the set $\{\nabla u_i(\boldsymbol{x}_j, t_k)\}_{ijk}$. A third-order $7 \times 7 \times 7$ Savitsky Golay filter was also applied to remove noise. The effects of the directional and smoothing filters are explored in Section 7.4. For now, this filtered displacement field is considered the 'true', noise-free one.

The three data sets mentioned in this section can now be used to compare the shear modulus inversion algorithms. The noise-free displacements are taken to be the two SG filtered simulated fields, and the directional/SG filtered measured field. To test the effect of noise, the simulated data sets are corrupted by AWGN, and the measured data set is used with no pre-processing. All three fields exist in the time-domain. To use them with the harmonic wave equation, they must be transformed to the frequency domain. It is expected that the first harmonic (equal to the driving frequency) contains most of the energy of the system, so this harmonic is used in the inversion procedures.

7.2 Accuracy

Before beginning any inversion procedure, the time-domain data is transformed into the frequency domain with the use of a Fast Fourier Transform (FFT), and the first harmonic is extracted. All first and second partial derivatives of the harmonic data are estimated using fourth-order $5 \times 5 \times 5$ Savitsky-Golay filters. This information is common to all methods. Accuracy is tested by comparing the resulting estimates with the 'ideal' shear modulus maps. The methods to be compared are AIDE, AOSMD, DIG_1–MW, and BIQO–E with overlapping blocks. Recall that AIDE is the original pointwise method, AOSMD is the approximate q-optimal Interior Helmholtz inversion, DIG_1–MW is the initial estimate using a Green's function and a moving window, and BIQO–E is a blockwise method that uses quadratic programming to estimate the shear modulus and its three first derivatives.

Inversions are performed for the three test-cases using the 'noise-free' displacement fields. The following method-specific parameters are implemented:

- AOSMD:
  - Window size: $5 \times 5 \times 5$
  - Derivative matrices: second-order finite difference, centred, with zero-extension assumption
- DIG_1–MW:
  - Window size: $5 \times 5 \times 5$
  - Window: W_{p7}
- BIQO–E:
  - Window size: $20 \times 20 \times 20$ (simulated), $20 \times 20 \times 14$ (measured)
  - Overlap: 50% with linear merge
  - Derivative matrices: fourth-order finite difference, centred, with up/down-winding on boundaries
  - Weight parameter: $\alpha = 1$
  - SVD tolerance: 1%

Unless otherwise specified, these parameters are used for all inversions in this chapter. The small window size for AOSMD and DIG was chosen in order to reduce the impact of the local homogeneity assumption. Different window sizes are tested in Section 7.2.1. Much larger windows are preferred in the BIQO methods in order to increase accuracy of the discrete derivative approximations. Since the measured displacement field only consists of 14 slices along the z-axis, the third window dimension is limited by this value. For these large sizes, it was found that the weight parameter, α, had little impact on the shear modulus estimate.

Table 7.1: Accuracy of inversion methods.

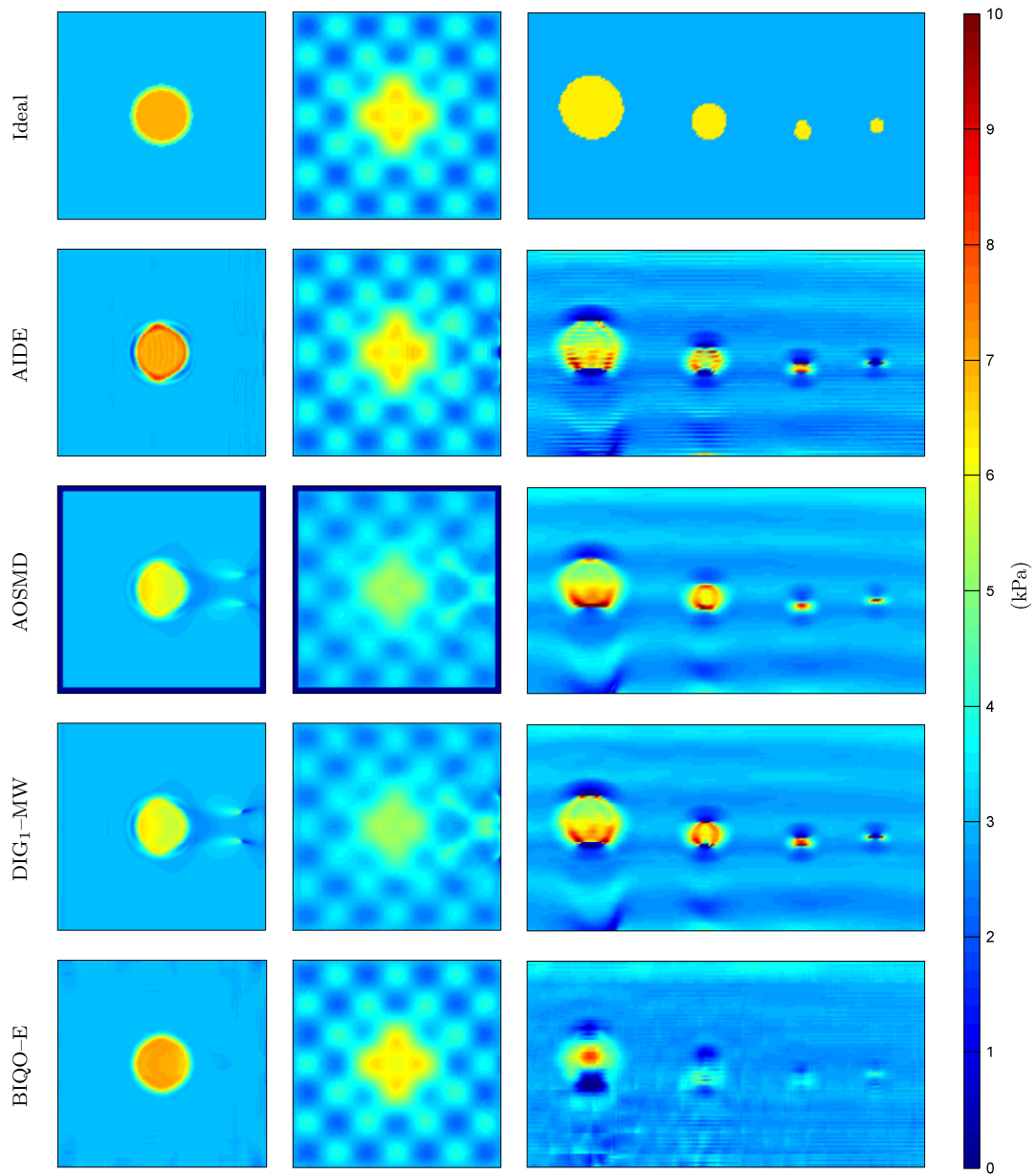| Method	Normalized Error (%)		
	Sim. #1	Sim. #2	Meas.
AIDE	12.51	8.81	31.62
AOSMD	11.87	11.14	46.73
DIG_1–MW	12.41	11.21	27.03
BIQO–E	3.42	1.26	31.00

Figure 7.3: Comparison of the accuracy for the three test-cases. Images depict the centre slice of the shear modulus estimates, parallel to the xy-plane.

Elastograms of the shear modulus estimates are shown in Figure 7.3. In order to quantitatively compare the estimates, the following normalized error norm is defined:

$$\text{Error}(\mathcal{M}) = \frac{\|\mathcal{M} - \mathcal{M}_{\text{ideal}}\|}{\|\mathcal{M}_{\text{ideal}}\|} \times 100\%.$$

The resulting errors for the four methods are given in Table 7.1. The relatively high error for the third test-case (measured) is expected because of the following:

· The original data is noisy.

· The 'ideal' shear modulus is estimated, based on expected moduli of the two composing gels.

· The discontinuity around the cylindrical inclusions invalidates the assumptions of the tissue model.

Still, all four estimates show evidence of the four cylinders.

The AIDE, AOSMD and DIG techniques all produce shear modulus estimates with similar shape features. In particular, there seems to be a skewing effect in the simulated test-cases. For these three methods, the local-homogeneity assumption is imposed, suggesting that this is the cause of the skewing behaviour. Quantitatively, it is difficult to compare these three methods. In the first case, the AOSMD is the more accurate of the three; in the second, AIDE is the more accurate; and for the measured data set, DIG is the more accurate.

In both the AOSMD and DIG estimates, the value within the spherical inclusion (sim. #1) is significantly underestimated. It is hypothesized that this is due to the decoupling assumption. This same behaviour was apparent in the BIQO–D example from the previous chapter. The features also seem to be blurred, which is particularly visible in the second test-case. This blurring is due to the averaging effect of the window, which is again a result of a homogeneity assumption. The ring around the boundaries of the AOSMD estimate should be ignored; the inversion was not performed in this region because the window would extend outside the domain. In the normalized error calculation, this boundary region is ignored.

The BIQO method is significantly more accurate than all other methods in the first two test-cases. This is likely because it does not require any homogeneity assumption, allowing the true shape of stiffness distribution to be captured in the simulated tissues. It is also the only method that accurately estimates the sharp transition surrounding the spherical inclusion. In the third test-case, however, there is a significant smoothing of the transitions. This seems to be a conflicting result. The cause of the smoothing is believed to be related to the data pre-processing, and will be examined further in Section 7.4. Quantitatively, it is outperformed by DIG_1–MW for the measured data set, but not significantly. Of the four methods, it is deemed to be the most accurate.
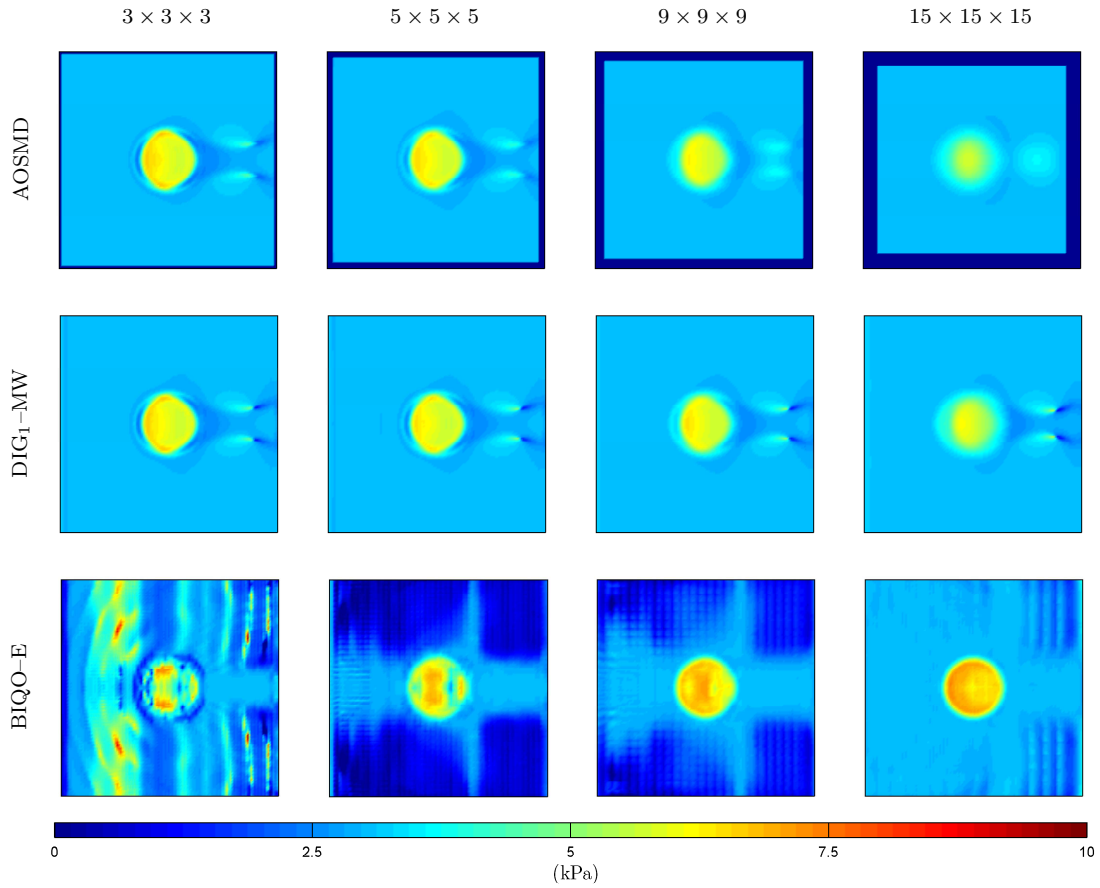
7.2.1   Window Size

The AOSMD, DIG and BIQO techniques all involve data windows. In AOSMD and DIG, these windows are used to reduce differentiation of the data and to regularize the solution. In BIQO, the window is used to separate the domain into manageable chunks. The size of these windows is likely effect the shear modulus reconstructions. This is explored here.

Figure 7.4 shows the shear modulus estimates using four different window sizes: $3 \times 3 \times 3$, $5 \times 5 \times 5$, $9 \times 9 \times 9$, and $15 \times 15 \times 15$. In order to implement the $3 \times 3 \times 3$ window, the derivative matrices in the

Table 7.2: Window size dependence.

Method	Normalized Error (%)			
	3^3	5^3	9^3	15^3
AOSMD	13.57	11.87	15.08	19.96
DIG_1–MW	13.71	12.41	14.30	17.07
BIQO–E	41.38	45.54	24.76	9.56

Figure 7.4: Comparison of the effect of window size on the shear modulus estimate. Inversions are performed for four different window sizes. Images depict the centre slice of the shear modulus estimate.

BIQO method were reduced to second-order. The normalized errors are shown in Table 7.2.

On the small 3^3 domain, the smooth tapering of the AOSMD and DIG windows is not well represented, even in the frequency domain. The tapered windows appear much like discrete δ functions, having a value of one in the centre, and zero everywhere else. When the support is increased slightly to $5 \times 5 \times 5$, the tapering behaviour is returned, and the accuracy is increased. After $5 \times 5 \times 5$, the accuracy of the two methods decreases as window size increases. This is largely due to the local homogeneity assumption. For these two methods, it is assumed that the shear modulus is approximately constant over the support of the window. In regions were the shear modulus changes, this assumption is invalid. As a result, regions of sharp variation are smoothed out. This is visible around the boundary of the spherical inclusion in Figure 7.4.

The BIQO method exhibits a different behaviour as window size increases. This is because there are no homogeneity assumptions imposed. Instead, as the window size increases, the discrete approximations

$$\frac{\partial \mathcal{M}}{\partial x} = D_1 \mathcal{M}, \qquad\qquad \frac{\partial \mathcal{M}}{\partial y} = D_2 \mathcal{M}, \qquad\qquad \frac{\partial \mathcal{M}}{\partial z} = D_3 \mathcal{M},$$

become more accurate. Here, D_i is a discrete approximation to the derivative operator acting along the ith direction. These conditions are used in a matrix inversion to find $\mathcal{M}$. Therefore, as the block-size increases, the accuracy of the shear modulus estimate within the block is also expected to increase. According to the quantitative measure of error, the 3^3 is estimate is more accurate than the 5^3, but this is not the case qualitatively; the 5^3 estimate clearly shows a stiff spherical inclusion in the centre, which is not very apparent in the 3^3 estimate. For larger windows, accuracy of the BIQO estimate increases as window size increases.

These comparison suggests that when there is no noise, small windows should be used for the AOSMD and DIG methods, and large windows for the BIQO method. The minimum size for AOSMD and DIG seems to be $5 \times 5 \times 5$, which is needed in order to represent the smooth tapering of the window. Odd lengths are required in order for the windows to possess a central point. For BIQO, the upper bound on block size is only limited by computational feasibility. Large blocks result in the need to solve large systems of equations. The size should be chosen to balance accuracy and speed. For the simulated data sets, a $20 \times 20 \times 20$ window is able to produce an estimate within 3.5% of the true shear modulus. Further enlarging of the window is not cost-effective; the marginal increase in accuracy is much less than the marginal increase in computation time.

7.3   Noise Sensitivity

The first data set, containing the spherical inclusion, is used to test the sensitivity to noise of the shear modulus reconstruction algorithms. The displacement field is corrupted with additive white Gaussian noise. The standard deviations of the noise is set to 0.5%, 1%, 2% and 5% of the forcing amplitude. The signal-to-noise ratios for each of the polarizations of motion are listed in Table 7.3. The SNR is calculated as follows:

$$\text{SNR}(u_i + \eta) = \frac{E\left\{u_i^2\right\}}{E\left\{\eta^2\right\}} = \frac{\int\limits_0^T \iiint\limits_\Omega \left[u_i(\boldsymbol{x}, t)\right]^2 \, d\boldsymbol{x} \, dt}{\sigma^2 T \iiint\limits_\Omega d\boldsymbol{x}},$$

where u_i is the time-dependent signal, and η is the noise. This is a ratio of the energies in the true signal and the noise. SNR is usually reported in decibels, which is calculated as $\mathrm{SNR_{dB}} = 10\log_{10}(\mathrm{SNR})$. Note that the displacements oriented in the x- and y-directions have a significantly lower SNRs than displacements oriented in the z-direction. Particularly, at 5% noise, the SNR for the y-displacements, in decibels, is negative. This means there is more energy in the noise than there is in the displacements. A slice of the field with the added noise is shown in Figure 7.5. Even at the 1% level, the x- and y-displacements are strongly corrupted by the noise. At 5%, the true displacements are unrecognizable. The results of the shear modulus estimates with varying noise levels are shown in Figure 7.6.

Table 7.3: SNR for simulated displacements.

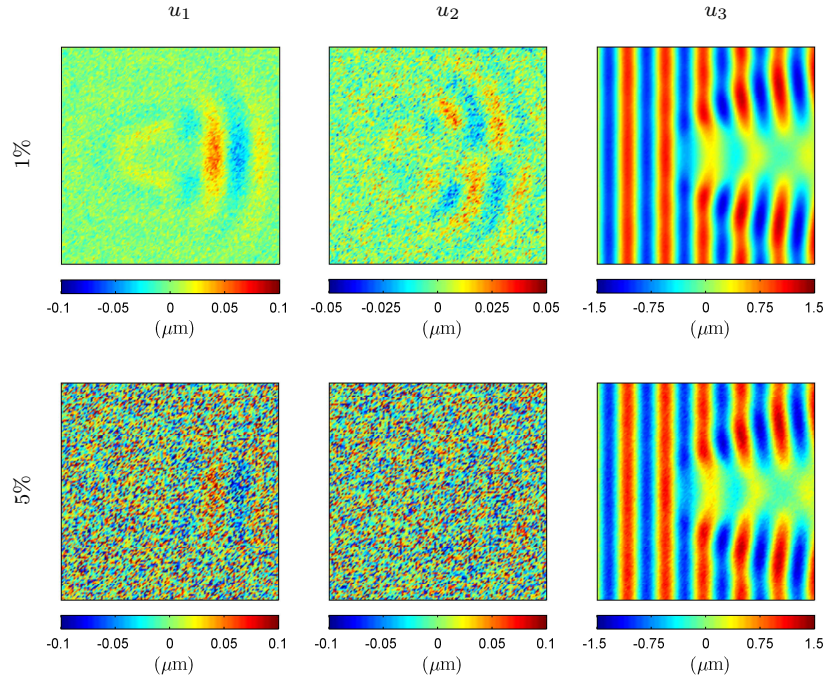| Polarization		SNR (dB)			
	$\sigma =$	0.5%	1%	2%	5%
u_1		22.28	16.26	10.24	2.28
u_2		17.55	11.53	5.51	-2.45
u_3		42.79	36.77	30.75	22.79

Figure 7.5: Slice of the displacement field with additive white Gaussian noise.

The AIDE and BIQO–E techniques are found to perform poorly in the presence of noise. These methods depend on the coupling between the equations of motion, so use all three polarizations. Even for small noise variances, the x and y polarizations are strongly corrupted. They have very small amplitudes relative to that of the tissue excitation, so are easily overpowered by the noise. The AIDE technique seems to out-perform BIQO for small noise levels. In AIDE, the shear modulus estimate at a single point is only affected by the noise in a small neighbourhood of that point. In BIQO, the estimate
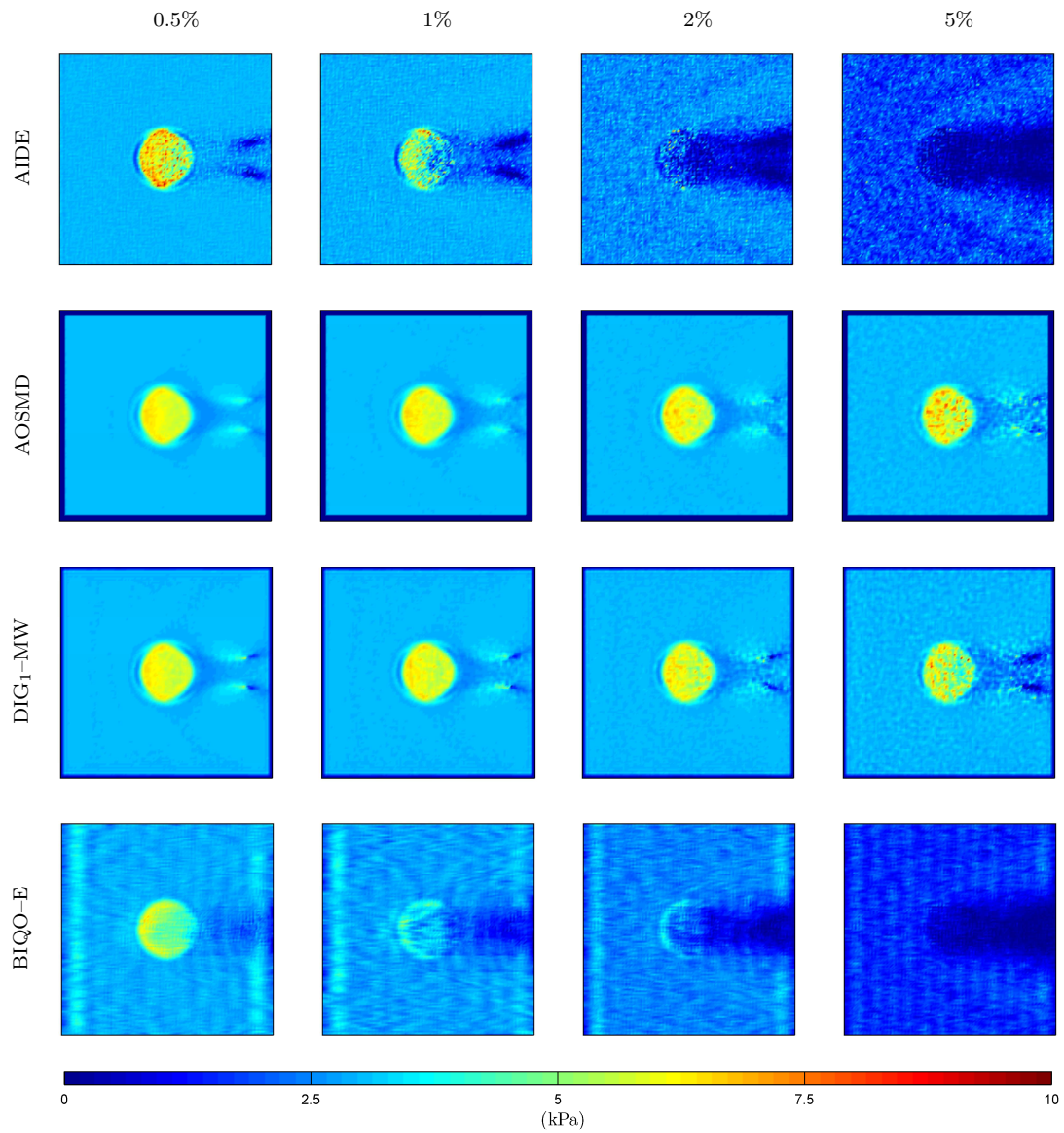
Figure 7.6: Comparison of the performance in the presence of noise.

of the shear modulus at a single point is dependent on data within the entire block. This means it is affected by the noise at all other points within the block. Reducing the block size might reduce the effect of the noise, but will also significantly reduce the accuracy of the method. It seems data pre-processing might be a necessary step if these methods are to be used in practice.

The AOSMD and DIG methods perform very well in the presence of noise, even at the 5% level. This is likely due to the following:

· The convolution reduces the need to differentiate noisy data.

· The regularization assumption, $\mathcal{M} \approx$ constant over window support, smooths the estimate.

· Decoupling allows the estimate to be obtained from a single polarization.

The AOSMD and DIG methods were designed to perform well in the presence of noise. It was discussed in Chapter 3 that differentiation of noisy data amplifies high frequency noise. Reducing the need for differentiation, therefore, will improve the robustness of the algorithm. In these methods, it is also assumed that $\mathcal{M}$ can be considered constant over small domains. The modulus estimate is effectively averaged over the support of the window, smoothing out noise. This, unfortunately, also smooths regions where the shear modulus is supposed to vary. Finally, the decoupling assumption allows an estimate of the shear modulus to be obtained from a single polarization. The displacements oriented in the same direction as the excitation (in this case, the z-direction) have a significantly larger amplitude than the other two polarizations. This leads to a much larger signal-to-noise ratio. This strongest polarization of motion is the only significant contributor to the shear modulus estimate in the AOSMD and DIG techniques. For these three reasons, the AOSMD and DIG methods are much less sensitive to noise than AIDE and IQO.

Table 7.4: SNR for noisy shear modulus estimates and two window sizes.

Method	SNR (dB)							
	$5 \times 5 \times 5$				$9 \times 9 \times 9$			
	$\sigma = 0.5\%$	1%	2%	5%	$\sigma = 0.5\%$	1%	2%	5%
AOSMD	44.60	38.41	32.55	24.63	59.71	54.02	47.96	39.94
DIG_1–MW	38.99	32.61	27.02	19.34	48.21	42.25	36.39	28.42

In order to compare the noise performance between the AOSMD and DIG methods, it is assumed that the noisy estimate of the shear modulus can be modelled as a random variable:

$$\hat{\mathcal{M}}(\sigma) = \mathcal{M} + \eta_\mu(\sigma),$$

where $\mathcal{M}$ is the noise-free estimate, and η_μ is a random noise parameter. From this model, a signal-to-noise ratio of the estimate can be defined:

$$\text{SNR}\left(\hat{\mathcal{M}}\right) = \frac{E\left\{\mathcal{M}^2\right\}}{E\left\{\left(\hat{\mathcal{M}} - \mathcal{M}\right)^2\right\}}.$$

This calculation is performed for both methods at each of the noise levels two window sizes. Results are shown in Table 7.4. In all cases, the AOSMD estimate has a larger output SNR than the DIG estimates.

This supports Oliphant's work in developing the 'optimal' AIDE technique [49]. There, estimators of the following form were considered:

$$\mathcal{M} = -\omega^2 \frac{h \star U}{\nabla^2 h \star U},$$

where $\star$ is used to represent a cross-correlation, and h is a filter with compact support. The optimal filters were defined as the ones that minimize the effect of the noise, so would therefore maximize the SNR of the shear modulus estimate. The filter used in AOSMD is one of these optimal filters. In chapter 5, it was shown that the DIG_1–MW estimator could also be written in the cross-correlation form. Therefore, its noise performance is bounded by that of the optimal method.

The sensitivity to noise of the AOSMD and DIG methods can be decreased by enlarging the window size. This improvement is apparent in Table 7.4. For a fixed level of noise, the larger window size has a larger shear modulus SNR. However, this comes at the cost of accuracy. In the AIDE and BIQO methods, there are no built-in features to handle noise. An improvement in noise performance can only be achieved with the use of pre-processing techniques.

7.4   Pre-processing

There is a common phrase used in the field of communication theory: *"Garbage in, garbage out."* If the input to a system is imperfect, then the output will be imperfect. For the MRE application, if the displacement data contain anything that is not captured by the model, this will corrupt the shear modulus estimate. The offending information is unwanted, and should be removed. This includes measurement noise, displacements caused by longitudinal waves, and internal reflections that might occur at discontinuous boundaries within the region of interest. Pre-processing is the act of attempting to remove this undesired information before the data is used in the main algorithm. In this section, three methods are compared: Savitsky-Golay smoothing, bandpass filtering, and directional filtering. Since the AOSMD and DIG methods are relatively insensitive to noise, the pre-processing methods are only tested for the BIQO–E method.

In Savitsky-Golay smoothing, a block of points is represented by a low-order polynomial. This limits the amount of allowable variation within the block, smoothing out the noise. In this section, a third-order SG filter with a block size of $7 \times 7 \times 7$ is used. A third-order polynomial is chosen because the displacement pattern in the tissue is expected to have inflection points. This behaviour requires at least third-order polynomials to accurately represent. The window size is kept relatively small in order to prevent over-smoothing, which might distort the shear waves.

For the bandpass filter, a third-order radial Butterworth filter, with cut-off frequencies of 5 and 250 cycles/m, is constructed. These parameters are chosen to closely match the filters used by Manduca *et al.* [39]. According to their work, removing the low spatial frequencies helps to remove displacements due to longitudinal waves and bulk effects. The high-frequency cut-off removes the high-frequency noise. These two frequency limits impose upper and lower bounds on the expected shear modulus. If the tissue is uniform and incompressible, then the spatial wavelength is related to the shear modulus,

$$\mathcal{M} = \frac{\rho \, \omega^2}{\|\boldsymbol{k}\|^2}.$$

This result was derived in Chapter 2. If $\|\boldsymbol{k}\|$ is restricted to be between 5 and 250 cycles/m, then

$$\mathcal{M} \in \left[\frac{\rho\,\omega^2}{250^2(2\pi)^2}, \; \frac{\rho\,\omega^2}{5^2(2\pi)^2} \right].$$

For $\rho = 1$ g/cm^3, and a forcing frequency of $\omega = 200\pi$ rad/s, the shear modulus estimate is approximately limited to the range $\mathcal{M} \in [0.16,\,400]$ kPa. All expected values of the shear modulus for the three examples tested here fall well within this range.

The bandpass filter can introduce artifacts near the boundaries of the domain if the data is not continuously periodic across the boundaries. This can be overcome by either continuously extending the data, or by using invertible tapering windows before filtering. For the tests in this section, the bandpass filter was applied to the data as-is, but only the central $80 \times 80 \times 20$ points were used for error calculations.

The final filter considered is the directional filter. Since the forcing was applied on the left-hand side of the tissue ($x = x_l$), most of the wave propagation is expected to be in the positive x-direction. Therefore, the filter was implemented with direction vector $\hat{\boldsymbol{n}} = (1, 0, 0)$. The tuning parameter $\alpha = 2$ was chosen, again to be consistent with Manduca *et al.* This directional filter is expected to remove all displacements caused by interior reflections, which might occur at the boundaries of inclusions. It also removes noise that is not seemingly propagating in the specified direction.

Table 7.5: Effects of pre-processing on BIQO–E estimate, spherical inclusion data.

Noise	Normalized Error (%)			
	SG	Band.	Dir.	B. & D.
0%	3.45	6.29	31.00	19.5
1%	8.06	11.31	19.77	14.05
5%	17.50	28.85	35.47	23.62

The three filters were applied to the first simulated data set with added noise. As a fourth test, the directional filter and radial bandpass filter were combined, approximately matching the pre-processing filter designed by Manduca *et al.* [39]. The centre slice of the BIQO–E estimates is shown in Figure 7.7. The normalized errors of the estimates for the 0%, 1% and 5% noise levels are given in Table 7.5.

In all three cases, SG filtering leads to the best improvement. The filtering technique does not significantly alter the noise-free solution, which is a desirable quality. The effect of the noise is almost eliminated at the 1% level. In fact, quantitatively, the BIQO method with 1% noise and SG filtering does quite a bit better than all the other methods applied to noise-free data. At the 5% level, the pre-processed estimate is comparable to the that with only 0.5% noise without pre-processing. This is a great improvement in performance.

The bandpass filter seems to be the next-best pre-processing tool. The exception is at the highest level of noise, where its performance is surpassed by the combination of the bandpass and directional filters. The shape of the inclusion is unchanged by the bandpass filter. The only significant change in the noise-free estimate is at the boundaries of the domain, which is a result of the displacements not being continuously periodic across these boundaries. This was accounted for in the normalized error calculation. The pre-processed estimate at the 5% level is comparable to that of the non-processed data at 1% noise, which is still a significant improvement.
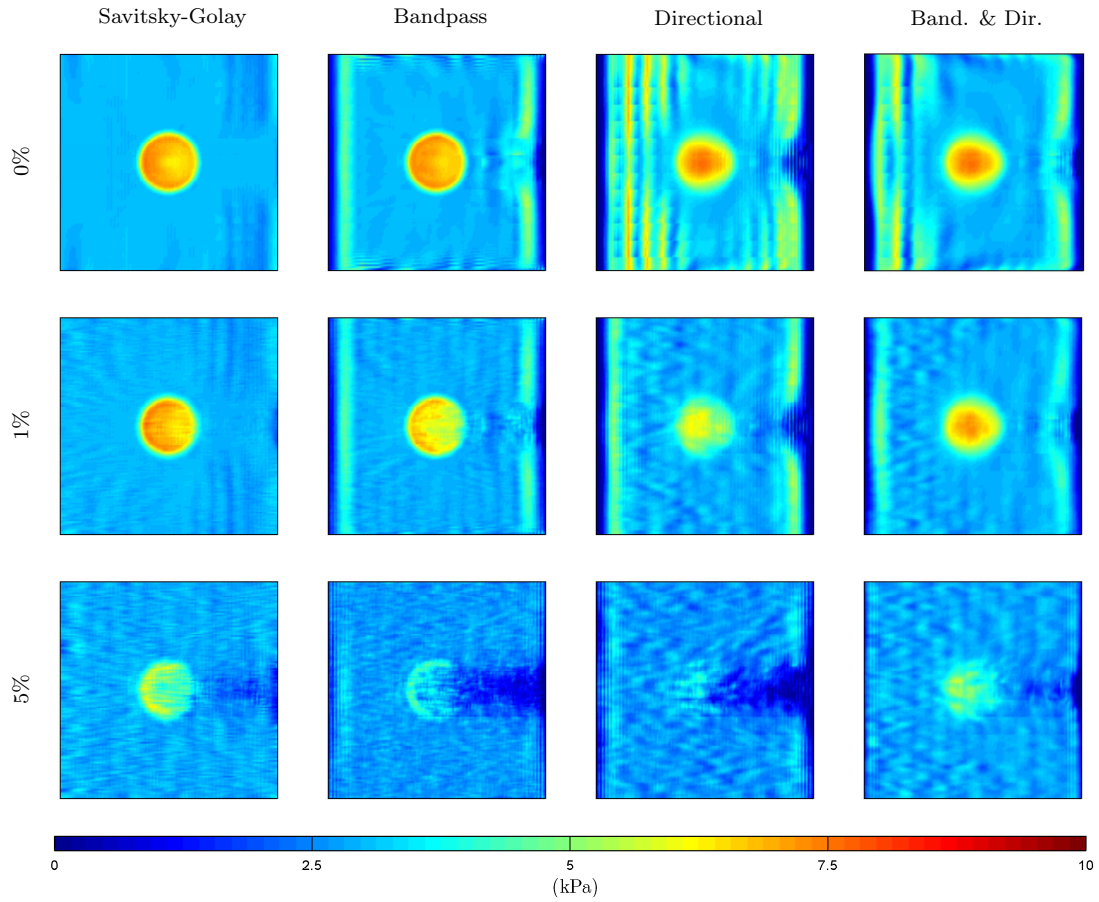
Figure 7.7: Comparison of pre-processing methods on the spherical inclusion data.

Table 7.6: Effective size of spatial support of the directional filter.

Tolerance	# Points	Dimensions
1%	339	77×9×3
2%	167	33×7×3
5%	65	13×5×1
10%	31	7×3×1

Unlike the previous two, the directional filter does not conserve the shape of the spherical inclusion. The sharp transition about the inclusion is smeared out. This same behaviour was apparent in the 'ideal' BIQO–E estimate using the measured data, back in Figure 7.3. The cause is the homogeneity assumption that must be made when the displacement field is replaced by the filtered version. This was discussed in Section 3.4.2. The amount of smearing is dependent on the size of the effective support of the filter. For a given tolerance, ϵ, the effective support of the directional filter is defined as the set of points satisfying:

$$\text{eff supp}(h, \epsilon) = \left\{ (\boldsymbol{x}, t) \mid \|h(\boldsymbol{x}, t)\| > \epsilon \max\left(\|h(\boldsymbol{x}, t)\|\right) \right\}, \tag{7.2}$$

where h is the impulse response of the filter. From this, an effective spatial support can be defined as the range of $\boldsymbol{x}$ within the effective support. For the directional filter considered here, the effective spatial support at several tolerance levels is listed in Table 7.6. Even at the high tolerance of 10%, the spatial support contains 31 points. In order for the filtered data to satisfy the original system of equations, the parameters in the system must be approximately constant over this effective support. A break in this assumption is expected to cause blurring or smearing. This is exactly what is observed.

The combination of directional and bandpass filters is able to remove more noise than either bandpass or directional filter alone. The bandpass filter will only eliminate noise frequencies outside of the pass-band. The directional filter is able to remove all noise that is not seemingly travelling in the direction of $\hat{\boldsymbol{n}}$. The term 'seemingly' is used because the random noise is not actually propagating; since the noise power is constant across all spatial and temporal frequencies, there is a component that will *appear* to move in any given direction. In the presence of high noise amplitudes, the benefit of removing more noise outweighs the smearing handicap, resulting in a more accurate estimate. Otherwise, the directional filter does not seem useful for the simulated data.

Table 7.7: Effects of pre-processing on BIQO–E estimate, cylindrical inclusion data.

	Normalized Error (%)		
	Unfiltered	Band-pass	Savitsky-Golay
Non-directional	44.03	39.75	36.62
Directional	40.45	40.72	31.00

The three pre-processing techniques were also tested on the measured data set. The shear modulus estimates and errors are shown in Figure 7.8 and Table 7.7, respectively. This time, the additional directional filtering step always seems to improve the BIQO estimates. The major difference between this example and the simulated ones is that there is a discontinuity in the shear modulus.

When the displacement field was simulated, a smooth transition was introduced around the inclusions. This ensures continuity of the shear modulus, preventing internal reflections. In practice, there may be regions of discontinuity. For the measured test-case, the shear modulus of the gel phantom has a discontinuous jump at the boundaries of the four cylinders. In Figure 7.9, the component of the displacement field removed by the directional filter is shown. Reflected waves are particularly visible in the z-polarization, u_3.

At the boundaries of the cylinders where the waves are reflected, the differential equations of motion do not hold. This seems to cause a ringing in the shear modulus estimate around the cylinders, seen as a sharp underestimate then overestimate across the discontinuity. This behaviour is not only apparent
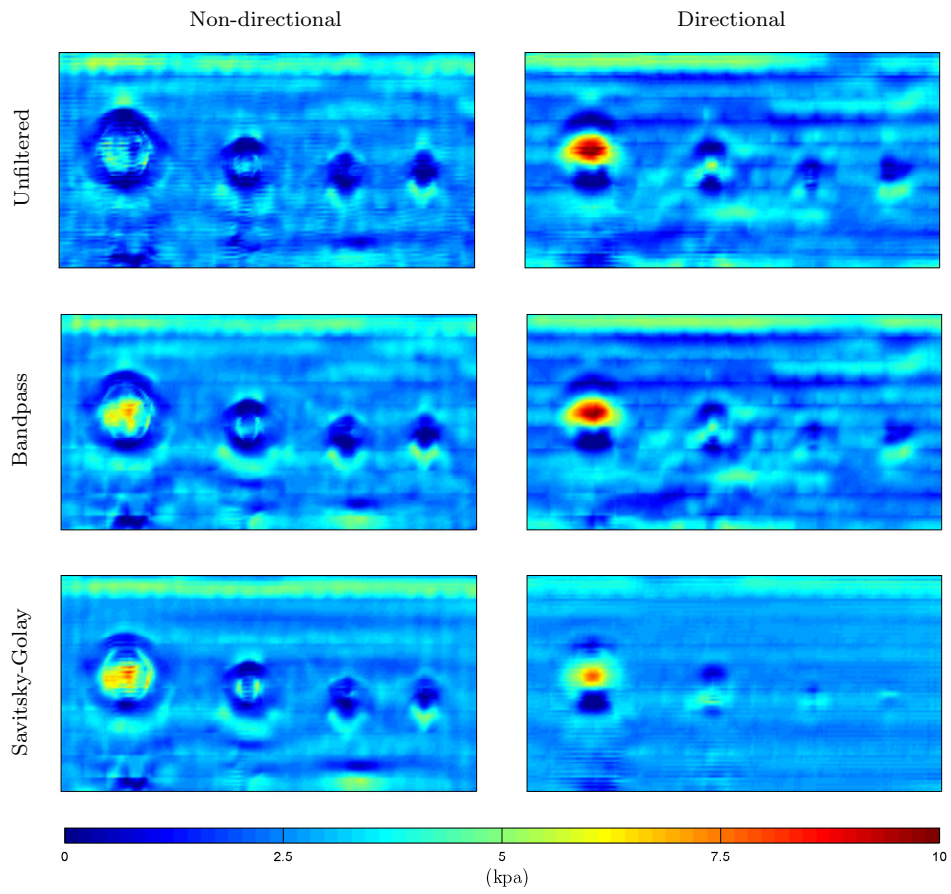
Figure 7.8: Comparison of pre-processing methods on the measured data.

(a) $(\delta - D_{\hat{\boldsymbol{n}}}^2) * u_1$　　　　　　(b) $(\delta - D_{\hat{\boldsymbol{n}}}^2) * u_2$　　　　　　(c) $(\delta - D_{\hat{\boldsymbol{n}}}^2) * u_3$
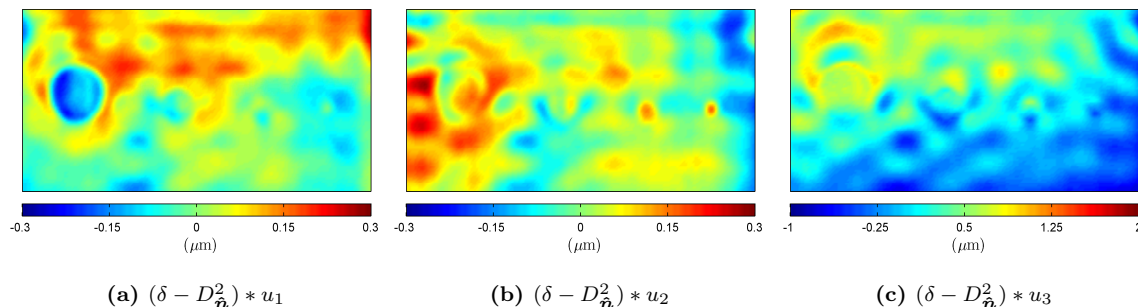
Figure 7.9: Component of displacements removed by the directional filter.

in the BIQO–E estimate, but also the AIDE, AOSMD and DIG methods, as shown in Figure 7.10. The directional filter removes this ringing. However, it cannot be concluded that the improvement is due to the removal of the reflected waves. The differential equations of motion are only invalid at the specific points where the reflections occur; the reflected waves, themselves, are harmless. Remember, the directional filter also imposes a regularity condition: the shear modulus must be approximately constant over the support of the filter. It is possible that the regularity condition is responsible for smoothing the estimate at the boundaries of the inclusions.

Table 7.8: Effect of directional filter on cylindrical inclusion data.

| | Normalized Error (%) | | | |
	AIDE	AOSMD	DIG$_1$–MW	BIQO–E
Non-directional	39.96	51.60	36.43	36.62
Directional	31.62	46.73	27.03	31.00

From these tests, it is not entirely clear whether directional filtering is necessary at the pre-processing stage. When the shear modulus is not constant, the waves are expected to travel in many directions, so directional filtering may be inadvisable. This was shown in the simulated fields back in Figure 7.1; the x- and y-polarizations contained waves travelling in many directions. The only case where directional filtering might help is if reflected waves cause destructive interference, resulting in low signal-to-noise ratios. However, this situation can usually be avoided by changing the location or frequency of the external forcing.

In this section, methods of pre-processing the data were explored in order to improve shear modulus estimation when the displacement field is imperfect. Pre-processing is not the only way to reduce the effect of noise. Averaging techniques performed during the inversion stage, like least-squares methods, can also help. The more data available, the more averaging can be done. This is the topic of the next section.

7.5　Multiple Scans

Inverse problems are usually difficult because of a lack of information. When this is the case, an inference must be made based on the little that is known. The more relevant information available,
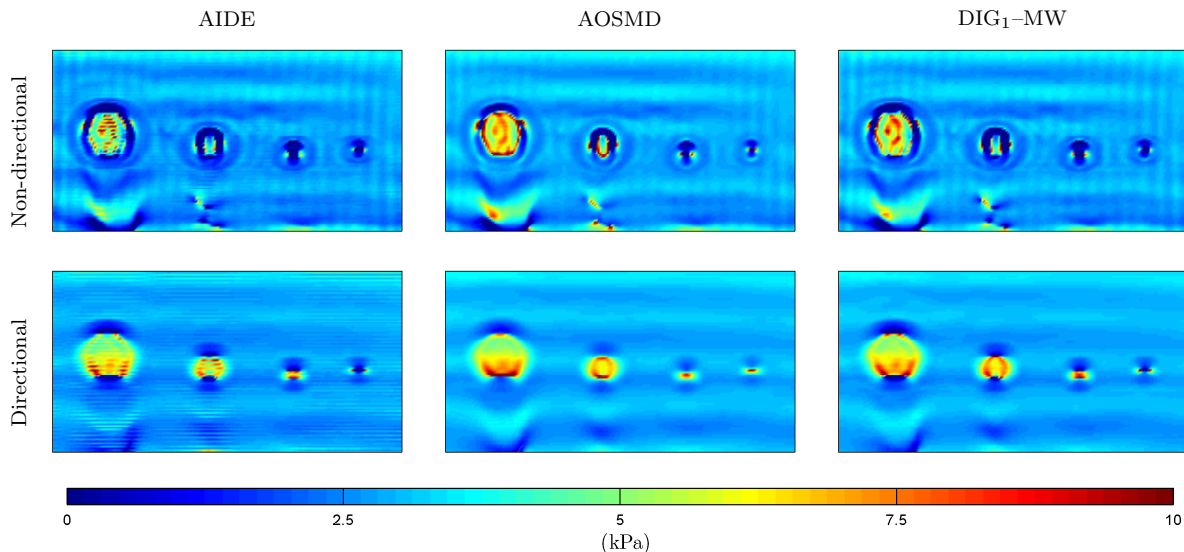
134

Figure 7.10: Effect of directional filter on the AIDE, AOSMD and DIG$_1$–MW estimates.

the more reliable the inference. For the type of differential equations involved in elastography, some form of external conditions are usually required in order to obtain a solution. Since the procedure is meant to be non-invasive, the stiffness at the boundaries of the region cannot be measured directly, so no boundary conditions can be provided. The only additional information available are displacement field measurements from repeated experiments.

Repeating the experiment has other benefits. When the tissue is excited from one side, any stiff inclusions can leave wakes of low amplitude displacements. This was visible in both the measured and simulated fields, shown in Section 7.1. Others have proposed the use of multiple mechanical drivers, exciting the tissue from multiple locations at once [31, 40, 41]. If forcing at multiple locations is done in separate scans, then further equations are generated, leading to a more determined system of equations. As long as the domains of these fields are spatially aligned, the information can be combined to solve for the common shear modulus. If the driver locations are chosen appropriately, areas of low SNR in one scan can have a higher SNR in another. In this way, the problem caused by wakes can be overcome.

Using multiple displacement fields can also improve performance in the presence of noise. Measurement noise between scans is expected to be uncorrelated. If the equations are combined and a least-squares method implemented, then some of the noise will be averaged-out.

For the IQO methods, there is an added benefit to using data from multiple displacement fields. The information from a second scan can be used to fill gaps in knowledge, such as the lack of boundary conditions. This is best illustrated by an example.

Example 6. *Recover the parameter, $\mathcal{M}$, from the following differential operator:*

$$L(\mathcal{M}, u) = \frac{d}{dx}\left(\mathcal{M}\frac{\partial}{\partial x}u\right),$$

135

where $\mathcal{M}(x) = 3 + \sin(x)$, using the two input functions: $u_1(x) = \sin(x)$, $u_2(x) = \cos(x)$.

This problem is modelled after the elasticity problem, with two measured displacement fields: u_1 and u_2. By evaluating $L(\mathcal{M}, u_i)$ for each of these, two differential equations are obtained:

$$\frac{d}{dx}[\mathcal{M}\cos(x)] = \cos(2x) - 3\sin(x), \tag{7.3a}$$

$$\frac{d}{dx}[\mathcal{M}\sin(x)] = \sin(2x) + 3\cos(x). \tag{7.3b}$$

No boundary information is given. In the AIDE, AOSMD and DIG techniques, regularizing assumptions are made for each equation in order to solve them separately. If there are multiple equations, then the separate solutions are combined with a least-squares method. Integrating the equations in (7.3) yields:

$$\hat{\mathcal{M}}\cos(x) = \sin(x)\cos(x) + 3\cos(x) + c_1,$$

$$\hat{\mathcal{M}}\sin(x) = \sin(x)\sin(x) + 3\sin(x) + c_2,$$

where $\hat{\mathcal{M}}$ is the regularized version of $\mathcal{M}$, and the constants c_1 and c_2 are determined by the regularizing assumptions. The least-squares solution for $\hat{\mathcal{M}}$ is

$$\hat{\mathcal{M}}(x) = 3 + \sin(x) + c_1\cos(x) + c_2\sin(x).$$

The presence of c_1 and c_2 implies that the final solution is dependent on the regularization used to find the individual solutions. If either of these constants are non-zero, the true parameter is not recovered.

In the IQO techniques, the system of equations is solved together. To proceed, expand the left-hand side of the equations in (7.3) using the product rule:

$$-\sin(x)\mathcal{M} + \cos(x)\frac{d\mathcal{M}}{dx} = \cos(2x) - 3\sin(x), \tag{7.4a}$$

$$\cos(x)\mathcal{M} + \sin(x)\frac{d\mathcal{M}}{dx} = \sin(2x) + 3\cos(x). \tag{7.4b}$$

This system can be written in matrix form and solved algebraically:

$$\begin{bmatrix} -\sin(x) & \cos(x) \\ \cos(x) & \sin(x) \end{bmatrix} \begin{bmatrix} \mathcal{M} \\ \frac{d}{dx}\mathcal{M} \end{bmatrix} = \begin{bmatrix} \cos(2x) - 3\sin(x) \\ \sin(2x) + 3\cos(x) \end{bmatrix}$$

$$\implies \begin{bmatrix} \mathcal{M} \\ \frac{d}{dx}\mathcal{M} \end{bmatrix} = \begin{bmatrix} 3 + \sin(x) \\ \cos(x) \end{bmatrix}.$$

With this method, the parameter and its derivative are perfectly recovered. The original system of equations has a completely determined solution; there is no need for regularization or external conditions. The additional differential equation, in some sense, accounts for missing boundary information.

A single differential equation does not have a unique solution without imposing external conditions, like boundary conditions. However, it does have a family of solutions, say $\mathcal{S}_1$. Two linearly independent differential equations have two different solution families: $\mathcal{S}_1$ and $\mathcal{S}_2$. If a function, s, must satisfy both

equations, then it must lie in the intersection of the two families: $s \in \mathcal{S}_1 \cap \mathcal{S}_2$. By adding more linearly independent equations, this intersection can be narrowed. For a system of N differential equations involving the single unknown function,

$$s \in \bigcap_{i=1}^{N} \mathcal{S}_i.$$

If the intersection consists of a single element, then a unique solution is obtained, even though no external conditions are imposed. This concept is elaborated in Appendix C. In the previous example, the two solution families intersect at the single function: $\mathcal{M} = 3 + \sin(x)$.

In the AIDE, AOSMD and DIG methods, a regularization condition is used to find a particular solution to each differential equation, separately. This means one element is chosen from each family of solutions, $s_i \in \mathcal{S}_i$. To obtain a single estimate for the entire system, a best-fitting one is constructed:

$$\hat{s} = \sum_{i=1}^{N} w_i s_i,$$

where w_i are weights, such as those that result from least-squares. There is no guarantee that $\hat{s}$ lies in the intersection of the solution families. It may not even lie in any individual family. This is a fundamental difference between the {AIDE, AOSMD, DIG} and IQO methods discussed in this dissertation, and explains why the IQO methods have a potential for higher accuracy.

7.5.1   Multi-Scan Inversion

For the two computer generated tissues, a set of linearly independent displacement fields can be obtained by repeating the forward simulations, applying the forcing terms at different boundaries. However, due to the rotational symmetries in the shear moduli and the grid, new displacement fields were obtained by simply rotating the original field about the z-axis by multiples of $90°$. In this way, four independent fields are generated for each of the two test-cases.

For the gel-phantom, only one displacement field was available. To test the algorithms using data from multiple scans, dummy-fields were created. The phantom, as a whole, is not reflectionally or rotationally symmetric. However, the inclusions themselves are. Four square prisms of the displacement field were extracted, each containing one of the inclusions at its centre. The shear modulus was estimated within each prism, separately. To generate multiple fields, the prisms of data were rotated about their centres at multiples of $90°$. This separation into prisms is depicted in Figure 7.11. The final modulus is constructed by simply re-inserting the four prism estimates. For the background, only two fields are used: the original, and one rotated by $180°$. The invalid portions in this background are over-written by the prisms. In this way, shear modulus estimates for up to four linearly independent displacement fields can be obtained.

The following rotations are used for the multi-scan inversions:

       2-Scan: $0°$, $180°$
       3-Scan: $0°$, $90°$, $180°$
       4-Scan: $0°$, $90°$, $180°$, $270°$

The shear modulus reconstructions using two linearly independent displacement fields are shown in
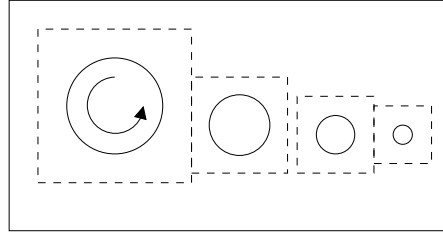
Figure 7.11: Dummy displacement field for the cylindrical inclusion data to test the use of multiple scans. The inclusions are separated into four square prisms. Each prism is rotated by multiples of 90° to generate up to four linearly independent displacement fields.

Figure 7.12. The errors for the 2-, 3-, and 4-scan inversions are given in Table 7.9. From the values in this table, it is apparent that performing a second scan can greatly improve the shear modulus estimate. This is true for all methods. Further scans do not lead to much more of an increase in accuracy. This is likely because regions with very small displacement amplitudes in one experiment have higher amplitudes when the forcing is applied to the opposite side of the tissue. Adding more than two fields does not drastically improve this SNR. However, additional fields can help remove directionally dependent distortions, such as those visible in the AIDE, AOSMD, and DIG estimates around the boundary of the spherical inclusion.

Table 7.9: Normalized error for an N-scan inversion.

| | | Normalized Error (%) | | | | |
		AIDE	AOSMD	DIG_1–MW	BIQO–E	PIQO
Sim. #1	1	12.41	11.87	12.41	3.42	–
	2	7.87	8.02	8.21	2.51	3.47
	3	7.50	7.85	7.97	2.26	3.24
	4	7.40	7.77	7.86	2.26	3.10
Sim. #2	1	8.81	11.14	11.21	1.26	–
	2	1.62	7.42	6.53	0.60	1.04
	3	1.30	7.39	6.49	0.54	0.95
	4	1.15	7.34	6.44	0.57	0.75
Meas.	1	31.62	46.73	27.03	31.00	–
	2	26.30	20.47	23.00	17.90	23.90
	3	26.47	20.53	21.85	18.55	23.48
	4	26.85	21.17	22.15	18.32	19.60

Since having information from two scans leads to more than four differential equations, the PIQO method could also be tested. With fewer than four equations, this method cannot provide a reliable estimate, as discussed in Chapter 6. PIQO is much more sensitive to noise than BIQO, which is why the estimate is less smooth and less accurate. However, it is a pointwise method, so is much faster. It also estimates all first derivatives of the shear modulus, which can help remove the noise and smooth the solution in a post-processing stage. In all three test-cases, the PIQO method has better performance than the other pointwise method, AIDE. The major difference between the two is that in AIDE, the
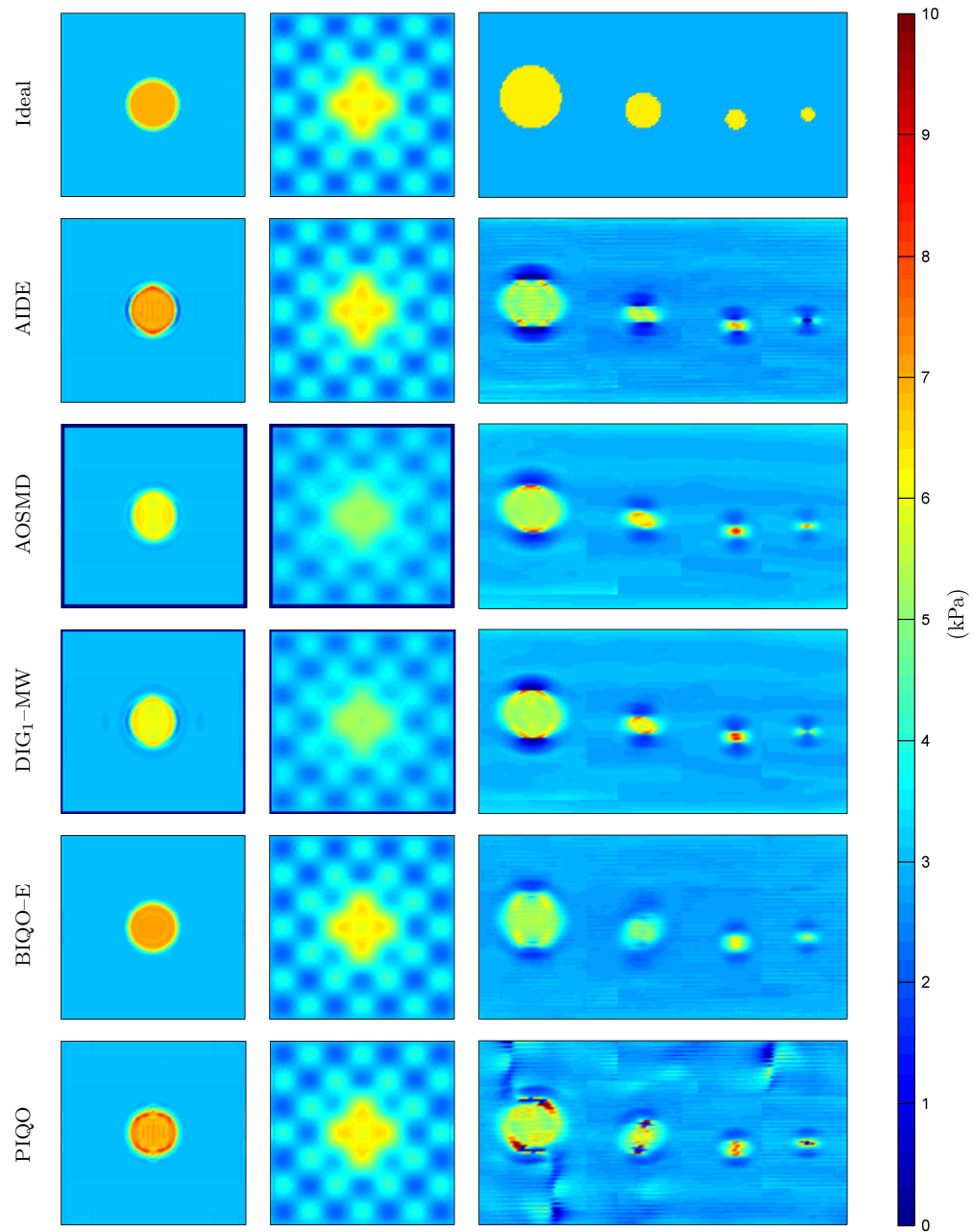
Figure 7.12: Shear modulus estimates using data from two scans.

gradient of $\mathcal{M}$ is assumed to be zero. In PIQO, $\nabla\mathcal{M}$ is estimated.

The three methods that rely on local homogeneity (AIDE, AOSMD, DIG) still show the sharp decrease then increase in the estimate around the boundaries of the inclusions, which is caused by the invalidity of the constant shear modulus assumption. This ringing behaviour is avoided in the IQO methods.

Quantitatively, BIQO is, without a doubt, the most accurate method when further scans are introduced. It is closely followed by the PIQO method. The other three methods are difficult to rank. AIDE is the next most accurate for the two simulated test-cases, but the least accurate for the measured data set. The DIG and AOSMD have comparable normalized errors for all three.

7.5.2   Multi-Polarization Inversion

Measuring the full three-dimensional displacement field requires repeatedly imaging the entire volume, while changing motion-sensitizing gradients. This can lead to long scanning times. If a reasonable estimate could be obtained using only one component of the field, then the total scan time can be drastically reduced.

By examining Figures 7.1 and 7.2, it is observed that the dominant component of the displacement field has the same orientation as the applied force. Only measuring this single polarization approximately reduces the total scanning time by a factor of three. In order to use a single component, it must be assumed that the equations of motion can be decoupled. This was done in the derivation of the decoupled AIDE (AIDE–D), AOSMD, DIG, and BIQO–D methods. To improve performance, the multi-scan approach can be applied. In this case, only one polarization of motion needs to be captured for each driver location.

Table 7.10: Normalized error for an N-polarization inversion.

		AIDE–D	AOSMD	DIG$_1$–MW	BIQO–D	PIQO–D
	Method \\ N	Normalized Error (%)
Sim. #1	1	10.06	9.55	10.26	110.21	–
	2	7.74	8.01	7.79	6.52	–
	3	7.37	7.84	7.50	6.15	–
	4	7.22	7.75	7.38	5.96	16.09
Sim. #2	1	7.67	8.38	7.89	49.43	–
	2	6.31	7.42	6.66	5.76	–
	3	6.27	7.39	6.65	5.62	–
	4	6.22	7.35	6.60	5.47	13.10
Meas.	1	37.09	32.94	42.23	289.31	–
	2	25.78	19.85	23.07	14.47	–
	3	24.99	19.78	20.95	15.02	–
	4	25.14	20.32	21.11	14.85	36.65

The shear modulus estimates obtained by AIDE–D, AOSMD, DIG$_1$–MW, and BIQO–D for two polarizations are shown in Figure 7.13. Since at least four equations are required for PIQO–D, the four-polarization estimate is given. Normalized errors are listed in Table 7.10.
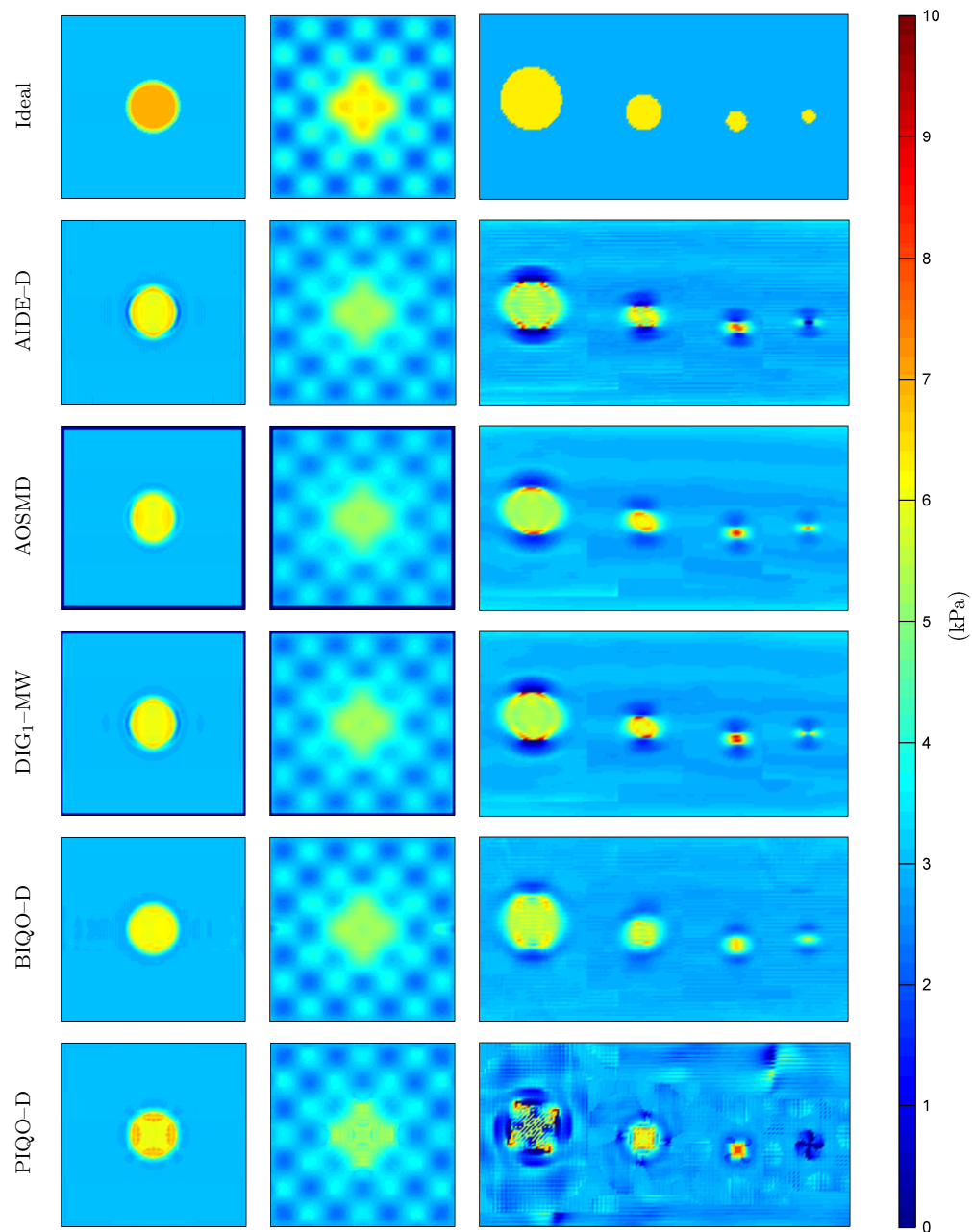
Figure 7.13: Shear modulus estimates using data from two polarizations for the AIDE, AOSMD, DIG and BIQO methods, and four polarizations for the PIQO method.

For a single polarization, only one differential equation involving $\mathcal{M}$ is generated. This means the solution is not fully determined without imposing external conditions. For the BIQO–D estimate, the magnitude of the gradient was minimized in order to regularize the solution. However, the method still attempts to estimate $\nabla\mathcal{M}$ based on the data, so is very sensitive to noise. This is the reason for the very large errors. When at least two polarizations are available, the BIQO–D method switches from being the least-accurate to the most-accurate. The reason is that the system becomes completely determined (at least in this numerical approximation), so no regularization is required.

It is again seen that a second scan leads to a substantial increase in accuracy for all methods. For more than two scans, the improvement is only marginal, if at all. The ringing behaviour is still evident in the homogeneity-based techniques. This time, all methods underestimate the values within the inclusions. This shows that the behaviour is, indeed, caused by the decoupling assumption, which was speculated in Section 7.2.

By examining Table 7.10, it can again be concluded that the BIQO method produces the most accurate shear modulus estimates when multiple polarizations are available. This time, the pointwise IQO method is the least accurate; there are not enough equations to sufficiently average out the noise. The other three methods have comparable accuracies for all three test-cases.
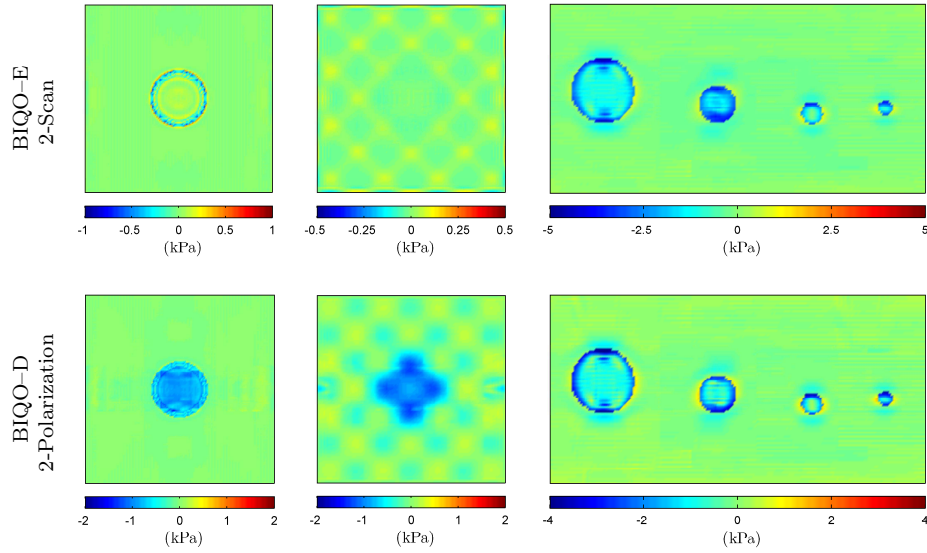

Figure 7.14: Plot of error, $\hat{\mathcal{M}} - \mathcal{M}$, between the BIQO-E/D estimates and the ideal solution.

Figure 7.14 shows the pointwise errors when using the full and decoupled versions of BIQO. When the full equation is used, the majority of the error for the first test-case is in the transition region from the background to the stiff inclusion. In the second-test case, the majority of the error occurs at the peaks and troughs of the background sinusoids. The cause of these errors is difficult to deduce, especially since the displacement fields are numerically generated. The differences might very well be caused by inaccuracies in the forward simulations. Regardless, these errors are very small compared to those observed when the decoupling assumption is made. With the decoupling assumption, the value within the spherical inclusion is underestimated by approximately 1 kPa. The same error is apparent in

the centre of the second test-case. This is evidence the the coupling terms are important when the shear modulus changes. In the third test-case, both methods are able to find the location of the cylindrical inclusions rather well. It is the radii that are underestimated. This may be due to the directional filtering, which is known to smooth the edges of sharp transitions. There is little difference between the BIQO–E and BIQO–D estimates for the gel phantom data.

Overall, the single-polarization inversions seem to give a good outline of the shear modulus distributions. Some of the values are underestimated, but this might not be too important if one is only searching for regions where the stiffness changes. If accuracy is important, then it seems the full equation of motion should be used, along with an inversion method that can utilize the coupling. Regardless, adding a second polarization of motion increases the accuracy of the estimates from all methods.

7.6 Summary

In this chapter four methods were analyzed and compared: AIDE, AOSMD, DIG$_1$–MW and BIQO–E.

The accuracy of the inversions was tested on three test-cases: two simulated tissues, and one gel phantom. Of the four methods, it was found that BIQO–E is the most accurate. The other three are difficult to rank; each seems to be slightly more accurate than the others for a different test-case. The AOSMD, DIG and BIQO methods all depend on window (or block) size. It was found that by increasing the window size, the AOSMD and DIG methods become less accurate, but the BIQO method becomes more accurate. In the first two methods, a homogeneity assumption is made over the support of the window. In regions where this assumption is invalid, there is a significant amount of smoothing, lowering the accuracy. This homogeneity assumption is not made in the BIQO method. Instead, increasing block size increases the accuracy of discrete derivative approximations, which allows for better results.

The four methods were also tested for robustness in the presence of noise. A noisy displacement field was created by adding white Gaussian noise with standard deviations equal to 0.5%, 1%, 2% and 5% of the forcing amplitude to one of the simulated data sets. The most robust method was found to be AOSMD. This is not surprising, since the method was designed for the specific purpose of minimizing the effect of noise on the estimate. The next most robust is DIG$_1$–MW, which uses a convolution to help average out noise, and is able to obtain an estimate from the polarization of motion with the highest signal-to-noise ratio. The AIDE and BIQO–E methods are very sensitive to noise. They depend on all three polarizations of motion. It was found that displacements oriented in the same direction as the external force have amplitudes an order of magnitude higher than the other orientations. Since the other two polarizations have such small amplitudes, they are heavily corrupted by even small amounts of noise. At the 5% noise level, it was even found that one of the polarizations contained less signal energy than the noise energy, making the signal impossible to use. If AIDE and BIQO are to be used in practice, a data pre-processing stage should be included.

Pre-processing is the act of removing unwanted information from a signal before the data is used. Three methods were compared in this chapter: Savitsky-Golay smoothing, bandpass filtering, and directional filtering. Of the three, SG smoothing led to the most improved shear modulus estimates, and did not distort the estimates when the data was noise-free. Bandpass filtering seems to be the next best pre-processing technique; in most cases it led to the second most improved estimates, and also did not cause any unreasonable distortions. The directional filter, on the other hand, did affect the estimates when the data was noise-free. This filter changes the shape of the underlying displacement field, making

it non-physical. This reshaped field will only satisfy the original equations of motion if the shear modulus is constant over the effective support of the filter, which was found to be quite large. The result is that variations in the shear modulus are smeared in the estimate. The directional filter did, however, seem useful for the gel phantom data. It removed the ringing effect from around the discontinuous internal boundaries of the cylindrical inclusions. At these boundaries, reflections are generated, invalidating the differential equations of motion (since the DuBois-Reymond lemma used in the derivations requires continuity). The directional filter removed these waves, which may explain why the ringing effect was removed. However, this cannot be concluded without further experiments. The removal of the ringing may also be a result of the homogeneity condition implicitly imposed.

Another way to improve performance is to repeat the experiment, changing the location of the external forcing. This generates a second, linearly independent, displacement field. The equations of motion can be viewed as a system of differential equations for the unknown shear modulus, $\mathcal{M}$, where the coefficients of the differential operator are directly related to the displacement field. Thus, adding further fields can result in a larger system of differential equations. With enough equations, this system becomes determined, eliminating the need for external conditions. Also, further equations can aid in the removal of noise. The underlying shear modulus is common to each repeated experiment. The noise, however, is expected to be uncorrelated and have a zero mean. Thus, by using more equations, the noise is expected to average out when a least-squares method is used. For all methods, adding the data from a second scan significantly improved the accuracy of the shear modulus estimates. Further repeated experiments only led to marginal gains, but did help remove directional artifacts.

Measuring the full displacement field for multiple experiments can be a very long process. Capturing only a single polarization of motion in each can greatly reduce scanning times. One polarization tends to dominate the field, anyway, suggesting a reasonable estimate might still be obtained. In order to use just one polarization, it must be assumed that the coupling terms in the equations of motion are negligible. This assumption was found to cause the shear modulus to be underestimated in the regions of high stiffness for the two simulated test-cases. Still, the overall pattern of elasticity was maintained, which might be all that is desired. The BIQO method was unable to accurately determine the shear modulus from a single equation; there are too many free parameters, and not enough data. As soon as a second polarization is added, however, it produced the most accurate estimate. In fact, for all methods and all test-cases, the addition of a second scan greatly increased the accuracy of the estimates. This is the same behaviour as in the multi-scan inversions, but it is now much more cost-effective to repeat the experiment.

Stiffness can be a strong indicator of pathological tissue. This has been known for thousands of years, making palpation one of the oldest diagnostic procedures. However, without exploratory surgery, a manual exploration of tissue by touch is limited to regions near the surface of the body. For this reason, the field of elastography was developed.

The stiffness of a tissue is a measure of how it responds to physical stress. This response can be measured by imaging techniques, non-invasively, allowing for a palpation-type diagnosis of tissues that are otherwise inaccessible. An advantage to this imaging procedure is that a quantitative measure of the stiffness can be obtained, allowing for statistical comparisons to aid in the classification of diseased tissue. Before such comparisons can be made, methods must be developed to accurately determine the true stiffness values. Unfortunately, imaging techniques cannot measure stiffness directly; they can only measure strains in response to known stresses. Mathematical models and methods are used to convert the strain response to a useable elastogram.

The central thesis of this dissertation is that fast, direct methods can be used to accurately reconstruct a quantitative measure of the stiffness in a tissue. It was assumed that the tissue can be modelled as an isotropic, linearly viscoelastic continuum. The goal was to develop more accurate methods of estimating the shear modulus distribution, by removing some of the common assumptions. Two new classes of methods were developed: Direct Inversion by Green's functions (DIG), and Inversion by Quadratic Optimization (IQO). These were compared to existing methods to see if a significant increase in accuracy can be achieved.

8.1 Summary

The key points from each chapter are highlighted here.

The various classes of elastography were introduced in Chapter 2. Since the methods developed in this thesis only apply to one class, focus was on direct methods for dynamic magnetic resonance elastography. A review of the relevant literature and an outline of the existing inversion techniques was provided. It was found that most methods assume the tissue is locally homogeneous, meaning its elasticity parameters are assumed constant over small domains. This assumption is counter-productive, since the goal of elastography is to find changes in stiffness. Another common assumption is that the three polarizations of motion are decoupled. This is based on local homogeneity and incompressibility. However, coupling is observed in practice. To accurately model the soft tissues, the coupling terms might need to be included. These are the two major assumptions that leave room for improvement in existing inversion methods.

In Chapter 3, all the relevant theory was provided that relates to the tissue model, the data acquisition procedure, inverse problems, and data processing. Each section contains important information that was used in the development of the direct inversion techniques that follow.

In the derivation of the tissue model, it was assumed that the material is isotropic and linearly-viscoelastic. It was also assumed that the motions are periodic in time, which led to the harmonic wave equations. In this model, there are two parameters of elasticity: the complex shear modulus, $\mathcal{M}$, and the complex first Lamé parameter, Λ. These parameters apply to general viscoelastic models, *but are dependent on frequency.* In order to remove this frequency dependence, a viscoelastic model must be specified. This means results of the elastograms performed at different forcing frequencies cannot be directly compared without a specific viscoelastic model. The final system of equations relates displacements, which can be measured, to two parameters of elasticity. The parameter of interest is the complex shear modulus. The other parameter presents a nuisance, and is usually ignored.

To acquire the data, it must be assumed that all motions are periodic with a known fundamental frequency. Imaging techniques cannot instantaneously measure full three-dimensional displacement fields. To capture a single 'snapshot' of the field in time, measurements must be made in batches at periodic intervals. Regular magnetic resonance imaging produces a density-type image. In order to depict motions, a special encoding sequence is used. The technique requires an added motion-encoding step, and uses phase-differencing. To obtain an image of a single polarization of motion, the entire volume must be scanned several times, modifying the motion-encoding gradient. Thus, capturing all three polarizations can be a long process. Repeating the experiment, changing the location of the external harmonic forcing, is very costly. This is why it is desirable to only require one polarization of motion.

Estimating the shear modulus from an equation of motion is an inverse problem. Such problems are typically ill-posed, which means the solution either does not formally exist, cannot be uniquely determined, or is unreliable due to over-sensitivity to small variations. When no solution exists, a best-fitting one can be constructed using least-squares methods. This idea is used in all inversion techniques when more than one polarization is considered. When the solution is not unique, the problem can be regularized by adding conditions. One way to accomplish this is to single out a particular solution that minimizes a quadratic functional. This led to Equality-constrained Quadratic Programming (EQP), and is the basis for the new IQO methods. Regularization is also required when the problem is too sensitive to small variations. Typical examples are ill-conditioned systems of equations. In such cases, the equations can be reduced to an 'effective' system, which is much more stable. System reduction can be performed using the described singular value decomposition technique. This reduced system can then be solved using EQP methods to steer any free parameters.

In the final section of Chapter 3, pre-processing techniques were discussed. Measured data is never perfect; it will always contain some form of unwanted information. The goal of pre-processing is to remove some of this information. Several filtering techniques were described, including Savitsky-Golay filtering, frequency band filtering, and directional filtering. Savitsky-Golay filters locally approximate a function with a low-order polynomial, and remove extraneous information. These are useful in removing noise and in estimating derivatives. In fact, SG filters were used for all methods in this dissertation to estimate derivatives of the data. Frequency band filters eliminate all information containing certain ranges of frequencies. One of the major problems with differentiating noisy data is that the high frequency components of the noise are amplified, and can drown out a signal. If the signal is known to have a finite bandwidth, then all frequencies outside this range can be attributed to noise and safely removed. Unlike the other filters, directional filters are not designed to eliminate noise. They are used to remove all information that is not propagating in a given direction. In MRE, this can help remove displacements caused by reflections. The filter changes the shape, or nature, of the data, making the displacement field

non-physical. It was shown that this new field does not satisfy the original equations of motion unless it is assumed the stiffness parameters are constant over the support of the filter. The effect of this was explored later in the analysis chapter.

The first detailed description of an inversion method was provided in Chapter 4. Algebraic Inversion of the Differential Equation (AIDE) is the basis of the new methods derived in this dissertation. In the original method, a local homogeneity assumption was made, which removed all terms involving derivatives of the shear modulus. The approximated system of equations can be considered algebraic in terms of $\mathcal{M}$, and was solved using least-squares. For further improvements, the system was reduced to an Interior Helmholtz equation, and re-posed in a variational form. Both sides of the Helmholtz equation were multiplied by a test function, integrated over the domain, and derivatives were transferred from the data to the test-function. To regularize, the shear modulus was assumed to be constant over the support of the test-function, so was taken outside of the integral. This allowed the estimate to be expressed as a ratio of two filtered versions of the displacements. 'Optimal' filters were defined as the ones that minimize the impact of noise on this estimate, and expressions for these were presented. The true optimal filters must be solved iteratively, but can be well-approximated by 'approximately-optimal' ones. One of these is the Approximately-Optimal, Smoothed and Matched Data (AOSMD) filter, which is used in the comparisons in Chapter 7.

Also in Chapter 4, some remarks are made about the results of the optimal AIDE technique. One of the purposes of introducing the filters was to eliminate the need to differentiate data. However, it happened that the optimal filter is a function of the data, so derivatives of the data were required anyway. Alternative arguments were also presented for why the optimal filters have a specific form. These arguments are based on much fewer assumptions than the original derivations.

In Chapter 5, the first novel technique was introduced: Direct Inversion by Green's functions (DIG). Like for the optimal AIDE, it was assumed the motions were uncoupled, and a local homogeneity assumption was made. This introduced a Laplacian operator. The Green's function for the Laplacian with no boundary conditions was used to try to solve this differential equation. Unfortunately, if no boundary conditions are applied, the solution depends on information outside the domain of measurement. To compensate for this, data windows were formally introduced into the system. Without further simplifications, an iterative technique is needed to solve for the shear modulus. As an initial estimate, it was assumed that the parameter was approximately constant over the support of the derivatives of the window. This can be slightly more relaxed than the optimal AIDE assumption if the window is constant on some interior domain. Several such windows were derived that use piecewise polynomials in transition regions. The assumptions are not well-founded when considering large domains; so, an overlapping block-decomposition method was proposed. The shear modulus was independently estimated on smaller blocks, and a global estimate was constructed by smoothly merging these block solutions. In the extreme case, when the blocks are shifted one point at a time, the result can be expressed as a convolution, which is much more computationally efficient to solve. This is the DIG_1–MW technique. It was shown that this estimate has the same form as Oliphant's optimal-AIDE, but there is a motivation for using a Green's function in the filter.

At the end of Chapter 5, a 'potential' function is introduced. The goal was to re-introduce the coupling terms, and remove part of the local homogeneity assumption. This work is preliminary. In testing on a sample data set, the potential was found to remove some artifacts, but introduced sharp spikes in the estimate, highly overestimating the shear modulus in some regions. There may be room

for improvement here.

Both the optimal AIDE and DIG techniques rely on some very limiting assumptions. The boundaries of pathological tissue, such as the boundary of a tumour, might be best marked by sharp changes in elasticity. When the local homogeneity assumption is made, these sharp changes are smoothed. Also, whenever the shear modulus is non-constant, or if the tissue is not perfectly incompressible, motions will be coupled. The best way to improve accuracy of the current methods is to remove these limiting assumptions. This was the goal of the IQO methods described in Chapter 6. Instead of performing regularization in the form of an assumption before the inversion, the regularization is performed *during* inversion, and *only if required*. To limit the number of free variables, the longitudinal component of the equation was still ignored. It was found that with at least four linearly independent differential equations, $\mathcal{M}$ and its three first derivatives could all be estimated as independent functions. This inversion is a pointwise operation (PIQO). Each full displacement field generates three equations, so it is enough to repeat the experiment once, changing the location of the applied force. This derivation treats the shear modulus and its derivatives as independent functions. However, there is a relationship between them. If these relationships are used, then fewer than four equations are required. In fact, it was found that in most cases, only two equations were required in order for the system to be fully determined. To implement the relationships, however, blocks of data must be considered at once (BIQO). Increasing block size was found to increase the accuracy of the inversion.

In the last chapter, four methods were compared: AIDE, AOSMD, DIG$_1$–MW, and BIQO–E. These comparisons were performed on three different data sets, two of which were simulated, and one that was obtained from an experiment performed on a gel phantom. Of the four inversion methods, BIQO–E was found to be the most accurate. The other three methods all had comparable accuracies, which is not so surprising, since they are all based on similar assumptions. To test performance in the presence of noise, one of the simulated data sets was corrupted with Additive White Gaussian Noise (AWGN) with standard deviations equal to 0.5%, 1%, 2% and 5% of the forcing amplitude. The AOSMD method was found to be the most robust, closely followed by DIG$_1$–MW. The other two methods performed very poorly in the presence of noise. To improve this performance, the data must be pre-processed.

Three pre-processing techniques were analyzed in Chapter 7: Savitsky-Golay smoothing, bandpass filtering and directional filtering. The first two methods attempt to remove noise. SG smoothing does so by approximating the data locally by low-order polynomials, and the bandpass filter removes noise by discarding all information containing frequencies outside of the pass-band. Of the two, the SG smoothing seemed to be the most effective. Directional filtering attenuates all waves not travelling in a particular direction. This can be used to eliminate waves caused by internal reflections, or separate complicated destructive interference patterns, leading to higher signal-to-noise ratios. For the simulated data, it was found that these filters do not help. In fact, the filters distort the shear modulus estimates, even when there is no noise. Directional filtering did, however, seem to increase accuracy for the measured data set. This may be due to the removal of reflected waves, or may be due to the implicit homogeneity assumption.

Performance was also shown to be improved when more than one displacement field is used for the inversion. Further linearly independent displacement fields can be obtained by repeating the experiment, changing the location of the actuators. As long as the domain of the second displacement field is aligned with the first, the spatial distribution of the shear modulus is in common. It was found that data from one additional scan significantly improved the accuracy of all methods. Further scans led only to

marginal gains, if any at all.

In the last section of Chapter 7, the decoupling assumption was tested. Measuring all three polarizations of motion is time consuming. It was found that the dominant component of the motion is oriented in the direction of the applied force, and has amplitudes an order of magnitude higher than motions in the other two orthogonal polarizations. Scan time can be reduced if only this one polarization is captured. For the simulated test cases, the decoupling assumption was found to cause an underestimation of the shear modulus values in the stiff centres. This effect is not a result of the value of the stiffness in the centre, but rather the change in stiffness as the waves propagate through the centre. By repeating the experiment with a different forcing location, this time only measuring the dominant polarization, accuracy was shown to increase for all methods. The final accuracy, however, is still limited by the validity of the decoupling assumption.

8.2   Conclusions

Of the two new approaches developed in this dissertation, the IQO methods have the greatest potential for increasing the accuracy of shear modulus estimates, as long as there is a high signal-to-noise ratio. The only advantage of the DIG methods is that they are fast. However, all other methods can produce estimates in a reasonable amount of time (up to an hour for the experiments considered in this dissertation), making this advantage insignificant. In the IQO methods, the local homogeneity assumption is removed, which allows the estimate to finally capture sharp changes in the shear modulus.

If the longitudinal component is ignored, it was found that the full system of differential equations *does* have a unique solution, $\mathcal{M}$, without the need for boundary or regularizing conditions. So, why do most groups make the local homogeneity assumption? Estimating the extra gradient term is not much more computationally challenging. Not only does this allow for local variations, but by not regularizing, one is more likely to obtain a solution that matches the 'true' stiffness. When regularization is applied, the validity of the estimate is dependent on the validity of the regularizing assumptions. Since the solution can be determined without regularization, there is no need for further assumptions that might corrupt the estimate.

Accuracy of the methods was also shown to greatly improve when data from more than one displacement field are used. Not only can this improve noise performance, but it can also lead to more determined systems of differential equations, allowing for a faster and more accurate estimation of the derivatives of $\mathcal{M}$. In practice, additional displacement fields from repeated experiments may first need to be spatially registered, but this is a well-studied problem in medical imaging. It is believed that there is a great potential for improvement in this area.

In conclusion, it was verified that direct methods are able to construct an accurate quantitative description of stiffness in a soft tissue. A class of methods, IQO, was created that do not rely on the local homogeneity assumption. These methods are not much more computationally challenging than the original AIDE technique, still use only local data, do not require any external conditions, but were found to be significantly more accurate. The development of these more accurate stiffness estimation techniques opens the door for improved quantitative analysis. This is one step in the right direction, perhaps leading the way for magnetic resonance elastography to become a reliable and commonly used diagnostic procedure.

FUTURE WORK

Elastography is still a relatively new field, and is not yet commonly used for diagnostics. Before the technique can be adopted, much more work is required to improve reliability and robustness. In this dissertation, the goal was to marginally increase the accuracy of inversion methods. Even in this small step, there are many ways in which the work contained in this document can be completed, extended or improved.

For the Direct Inversion by Green's function method, an expression was derived that includes the shear modulus in an integration term. It was stated that an iterative technique needs to be developed to solve this. One can begin by assuming the shear modulus is constant around the boundaries, as was done in this thesis. Once an initial estimate is obtained, the values determined along the boundaries can be used as boundary conditions for the next iteration. In this way, the true boundary values will hopefully emerge.

Rather than impose the local homogeneity assumption in order to introduce a Laplacian operator, one can attempt to numerically estimate the true Green's function. This function will depend on the data, and some form of boundary conditions will need to be imposed. Some regularization might be required, since the Green's function will be affected by noise in the measured data. The boundary conditions may be adapted, iteratively, to converge to a best-fitting solution. By removing the local homogeneity assumption here, it is hoped that sharp transitions in the shear modulus can be captured accurately, like in the IQO techniques.

Another possible direction is to explore the use of the potential function in Section 5.2. This may have the ability to re-introduce the coupling terms and allow for local variation, while still taking advantages of the noise performance of the DIG or optimal AIDE techniques. Preliminary tests have shown that some wave-like artifacts can be removed when the shear modulus is slowly varying. However, in regions where there are sharp variations, this potential seems to over-sharpen the estimate, and produces large peaks. By exploring the cause of this, one might be able to improve the method.

In the BIQO method, for large block sizes and cases where the system is underdetermined, the standard null-space and range-space methods are far too slow to be considered practical. This is mainly because both were implemented using singular value decompositions, which is computationally expensive for very large systems. Faster, perhaps iterative, techniques for performing the quadratic optimization should be explored.

One of the disadvantages of the IQO techniques is that they are very sensitive to noise. This can be improved by pre-processing the data, but there may be a way to improve the method, itself. Finite differencing matrices may not be the best way to represent the relationships between the shear modulus and its derivatives. Perhaps smoothed derivative estimates, like those from Savitsky-Golay filters, are more appropriate and will lead to better performance in the presence of noise.

Most of the methods discussed ignore the longitudinal component of the wave. Part of the reason for

this is that the waves travel so fast that they cannot be accurately captured by the imaging procedure. However, they still do have an affect on the equations of motion. Inversions might be improved by including a term to account for this, and leaving it as a parameter to be estimated, like a pressure term.

The idea of using repeated experiments seems to show a lot of promise. That is, if it is practical in the lab. The separate experiments might be performed in one sitting, with multiple, but separate, drivers attached to the tissue, and only one activated per scan. If the driver needs to be repositioned between experiments, the tissue will likely have adjusted positions. In this case, the inversion methods will need to be combined with image registration techniques in order to align the two fields. Obtaining more data opens the door for estimating more parameters. For example, the density can be re-included, the longitudinal component or pressure term, perhaps even non-isotropic parameters can be introduced into the model and estimated. The resulting equations will still form a differential system with coefficients dependent on the displacement field, which means obtaining further fields can lead to larger systems of differential equations.

Another area that requires further investigation is the theory behind uniqueness. Without any external conditions, under what circumstances will the differential system of equations have a unique shear modulus solution? Some ideas are explored in Appendix C. From the simulations performed in this dissertation, it seems that there is always a unique solution when only two equations are used. This might be a result of the discretization of the problem, or may be a result that can be proved for the specific equations used in dynamic elastography.

One of the largest problems found in this dissertation is resolving sharp transitions in the shear modulus. Even the IQO methods were unable to resolve the boundaries of the cylindrical inclusions in the measured data set. When there are discontinuities, the differential equations of motion are no longer valid. Instead, the integral version must be used. One possible direction is to develop inversion methods that use this integral equation of motion.

GREEN'S FUNCTIONS

Green's functions were originally developed by a British mathematician in the 1830's, George Green. They are used extensively in the study of partial differential equations, and are extremely useful in solving inhomogeneous differential equations that are subject to known boundary conditions. For a linear differential operator, L, acting on the space of distributions, the Green's function of the operator satisfies:

$$L\mathcal{G}(\boldsymbol{x}, \boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}),$$

where δ is the Dirac delta distribution, and L only acts on the variable $\boldsymbol{x}$. Now suppose one wishes to solve an inhomogeneous problem of the type:

$$Lu(\boldsymbol{x}) = f(\boldsymbol{x}).$$

Using properties of the Dirac delta, and integrating over the $\boldsymbol{y}$ variable,

$$
\begin{aligned}
f(\boldsymbol{x}) &= \int \delta(\boldsymbol{y} - \boldsymbol{x}) f(\boldsymbol{y}) \, d\boldsymbol{y} \\
&= \int L\mathcal{G}(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \, d\boldsymbol{y} \\
&= L\left(\int \mathcal{G}(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \, d\boldsymbol{y} \right).
\end{aligned}
$$

This expression can be substituted into the original differential equation:

$$
\begin{aligned}
Lu(\boldsymbol{x}) &= L\left(\int \mathcal{G}(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \, d\boldsymbol{y} \right) \\
\implies \quad u(\boldsymbol{x}) &= \int \mathcal{G}(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \, d\boldsymbol{y} + h(\boldsymbol{x}),
\end{aligned}
$$

where h is a solution to the homogeneous problem $Lh = 0$, and accounts for the boundary conditions on u. In this way, a particular solution to the differential equation can be found. If the Green's function is designed so that the particular solution already satisfies the boundary conditions, then $h = 0$.

Not all differential operators have a Green's function. Unfortunately, the divergence operator, which appears in the equations of motion for elastography, is one of these. Instead, the Green's function for the Laplacian operator is considered.

A.1 Derivation of the Green's Function

Theorem A.1. *The free-space Green's function for the three-dimensional Laplacian operator is given by:*

$$\mathcal{G}(\boldsymbol{x}) = -\frac{1}{4\pi\|\boldsymbol{x}\|}.$$

Proof. The Green's function for the Laplacian operator satisfies the differential equation:

$$\nabla^2 \mathcal{G} = \delta.$$

The δ function is spherically symmetric, which suggests that the Green's function is also spherically symmetric. Using this symmetry, the differential equation can be re-written in spherical coordinates:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \mathcal{G}(r)}{\partial r}\right) = \delta(r).$$

Performing the integrations in terms of r,

$$\begin{aligned}
\mathcal{G}(r) &= \int\left[\frac{1}{r^2}\int r^2\delta(r)\,dr\right]dr \\
&= -\frac{c_1}{r} + c_2,
\end{aligned}$$

where c_1 and c_2 are constants of integration. Since this Green's function is to be used in a convolution, $\mathcal{G}$ must be square-integrable. Otherwise, $\mathcal{G} * f$ will not exist for all $f \in C^0\left(\mathbb{R}^3\right)$. This forces the constant c_2 to be zero.

The Green's function must also satisfy the Divergence Theorem, which states:

$$\iiint_\Omega \nabla\cdot\nabla\mathcal{G}\,dV = \iint_{\partial\Omega}\nabla\mathcal{G}\cdot\hat{\boldsymbol{n}}\,dA,$$

for an arbitrary volume Ω, its boundary surface $\partial\Omega$, and unit surface normal $\hat{\boldsymbol{n}}$. This relation will determine a unique constant c_1. For simplicity of calculation, take Ω to be a sphere of radius R. The left-hand integral evaluates to one for any volume containing the origin. The right-hand integral becomes

$$\begin{aligned}
\iint_{\partial\Omega}\nabla\mathcal{G}(R)\cdot\hat{\boldsymbol{r}}\,dA &= \iint_{\partial\Omega}\frac{\partial\mathcal{G}(R)}{\partial R}\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{r}}\,dA \\
&= \frac{c_1}{R^2}\iint_{\partial\Omega}dA \\
&= \frac{c_1}{R^2}\left(4\pi R^2\right).
\end{aligned}$$

Equating the two integrals gives $c_1 = 1/4\pi$, and the result follows. $\qquad\square$

A.2   Approximating the Green's Function

Since the measured data is in a discrete form, a convolution with the Green's function must be discretely approximated. The standard Riemann sum approximation fails, since $\mathcal{G}$ has a singularity at the origin. However, even though $\mathcal{G}(0)$ is undefined, the integral of $\mathcal{G}$ over a finite volume containing the origin *is* defined:

$$\iiint\limits_{V} |\mathcal{G}(\boldsymbol{x})|\, dV \leq \iiint\limits_{B_r} |\mathcal{G}(\boldsymbol{x})|\, dV$$

$$= \iiint\limits_{B_r} \frac{1}{4\pi \|\boldsymbol{x}\|}\, dV$$

$$= \frac{r^2}{2} < \infty,$$

where B_r is a ball of radius r containing the original volume (i.e. $V \subset B_r$). If it is assumed that f can be approximated by piecewise constants over its domain Ω, then

$$\mathcal{G} * f = \iiint\limits_{\Omega} \mathcal{G}(x - \alpha, y - \beta, z - \gamma)\, f(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma$$

$$\approx \sum_k \sum_j \sum_i f(x_i, y_j, z_k) \int\limits_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int\limits_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \int\limits_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \mathcal{G}(x - x_i - \alpha, y - y_i - \beta, z - z_i - \gamma)\, d\alpha\, d\beta\, d\gamma,$$

where i, j and k are indices that run over all grid points in Ω. In this manner, the continuous convolution can be approximated by the discrete convolution,

$$\mathcal{G} * f \approx \hat{f} * \hat{\mathcal{G}},$$

where $\hat{f}_{ijk}$ are the discrete measurements of f at grid points (x_i, y_j, z_k), and $\hat{\mathcal{G}}$ is the discrete Green's filter with entries defined as:

$$\hat{\mathcal{G}}_{ijk} = \int\limits_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int\limits_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \int\limits_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \mathcal{G}(x_i + \alpha, y_i + \beta, z_i + \gamma)\, d\alpha\, d\beta\, d\gamma.$$

This Green's filter is defined everywhere. Since it only depends on the grid spacings, the filter can be pre-computed and stored. Evaluation of this integral is accomplished using an adaptive Simpson quadrature technique.

A.3   Approximating Derivatives of the Green's Function

In the MRE application, data is only known over a finite domain. The Green's function derived in this chapter does not take the boundary of the domain into consideration. To handle handle this, window functions were introduced in Chapter 5. As long as the support of the window function is inside the

domain of measurement,

$$\mathcal{G} * \left(W\nabla^2 f\right) = Wf - \underbrace{\mathcal{G} * \left(2\nabla W \cdot \nabla f + f\nabla^2 W\right)}_{\mathcal{B}}. \tag{A.1}$$

The left-hand side requires second derivatives of f, but the right-hand side only requires first derivatives. However, derivatives of the window must be known. When the standard boxcar window is used, δ and δ' distributions appear in ∇W and $\nabla^2 W$ due to the discontinuities. When the convolution is applied, the boundary term becomes

$$\begin{aligned}
\mathcal{B}(\boldsymbol{x}) = &\int_\Omega \mathcal{G}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\alpha - x_l) - \delta(\alpha - x_r)\right] \frac{\partial f}{\partial \alpha}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} + \int_\Omega \mathcal{G}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\beta - y_l) - \delta(\beta - y_r)\right] \frac{\partial f}{\partial \beta}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \\
&+ \int_\Omega \mathcal{G}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\gamma - z_l) - \delta(\gamma - z_r)\right] \frac{\partial f}{\partial \gamma}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} - \int_\Omega \frac{\partial \mathcal{G}}{\partial x}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\alpha - x_l) - \delta(\alpha - x_r)\right] f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \\
&- \int_\Omega \frac{\partial \mathcal{G}}{\partial y}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\beta - y_l) - \delta(\beta - y_r)\right] f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} - \int_\Omega \frac{\partial \mathcal{G}}{\partial z}(\boldsymbol{x} - \boldsymbol{\alpha}) \left[\delta(\gamma - z_l) - \delta(\gamma - z_r)\right] f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha},
\end{aligned} \tag{A.2}$$

where $\boldsymbol{\alpha} = (\alpha, \beta, \gamma)$, $\boldsymbol{x} = (x, y, z)$, and Ω is the support of the window. This expression requires the derivatives of $\mathcal{G}$ to be approximated. Since $\mathcal{G}$ has a pole at the origin, its derivatives have a pole of even higher order at the origin. One must ensure that the approximated versions of these derivatives are still well-defined.

Consider one of the potentially undefined terms:

$$\int_\Omega \frac{\partial \mathcal{G}}{\partial x}(\boldsymbol{x} - \boldsymbol{\alpha}) \, \delta(\alpha - x_l) f(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha}.$$

Since f is only known at discrete points, it is assumed to be composed of piecewise constants. This allows the integral to be split,

$$\int_\Omega \frac{\partial \mathcal{G}}{\partial \alpha}(\boldsymbol{\alpha} - \boldsymbol{x}) \, \delta(\alpha - x_l) f(\boldsymbol{x}) \, d\boldsymbol{x}$$

$$\approx \sum_k \sum_j \sum_i f(x_i, y_j, z_k) \int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \frac{\partial}{\partial \alpha} \mathcal{G}(x - x_i - \alpha, y - y_i - \beta, z - z_i - \gamma) \, \delta(\alpha + x_i - x_l) \, d\alpha \, d\beta \, d\gamma.$$

When this is evaluated at all grid points, the only terms that might be undefined are when $x = x_i$, $y = y_j$ and $z = z_k$. At these points, the integral component becomes

$$\int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \frac{\partial}{\partial \alpha} \mathcal{G}(-\alpha, -\beta, -\gamma) \, \delta(\alpha + x_i - x_l) \, d\alpha \, d\beta \, d\gamma$$

$$= \frac{1}{4\pi} \int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \frac{\alpha}{(\alpha^2 + \beta^2 + \gamma^2)^{3/2}} \, \delta(\alpha + x_i - x_l) \, d\alpha \, d\beta \, d\gamma,$$

which is only non-zero for $|x_i - x_l| < \Delta x$. Thus, it must be shown that

$$\left| \frac{1}{4\pi} \int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \frac{x_l - x_i}{\left((x_l - x_i)^2 + \beta^2 + \gamma^2\right)^{3/2}} \, d\beta \, d\gamma \right| < \infty. \tag{A.3}$$
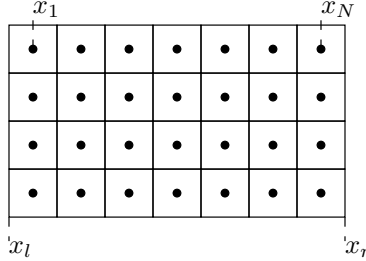

Figure A.1: Grid layout. The grid points lie in the center of grid spaces.

In practice, the grid points are assumed to lie at the center of grid cubes, so no points are exactly on the boundary; there is a shift by half a grid spacing: $x_1 - x_l = \frac{1}{2}\Delta x$. This is depicted in Figure A.1. Thus, the denominator in the above integral is never zero. However, one may be interested in the behaviour as $x_i \to x_l^+$. Instead of integrating over a square, consider integrating over a circle of radius R:

$$\frac{1}{4\pi} \int_0^{2\pi} \int_0^R \frac{x_l - x_i}{\left((x_l - x_i)^2 + r^2\right)^{3/2}} r \, dr \, d\theta = \frac{x_l - x_i}{2} \left[\frac{1}{x_i - x_l} - \frac{1}{\sqrt{(x_l - x_i)^2 + R^2}} \right]$$

Taking the limit as the grid point approaches the boundary,

$$\lim_{x_i \to x_l^+} \frac{x_l - x_i}{2} \left[\frac{1}{x_i - x_l} - \frac{1}{\sqrt{(x_l - x_i)^2 + R^2}} \right] = -\frac{1}{2}.$$

This limit is independent of the radius. Since the integrand in Equation (A.3) is strictly negative, there are two radii, R_1 and R_2, such that the integral over the rectangular domain is bounded between the two integrals over the circular domains, B_{R_1} and B_{R_2}. Therefore,

$$\lim_{x_i \to x_l^+} \frac{1}{4\pi} \int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \frac{x_l - x_i}{\left((x_l - x_i)^2 + \beta^2 + \gamma^2\right)^{3/2}} \, d\beta \, d\gamma = -\frac{1}{2}.$$

This means the boundary term, $\mathcal{B}$, is well-defined for all points. The final result is that the approximated Green's function and derivatives can be used to invert the windowed Laplacian in Equation (A.1).

Example 7. *Using the approximated Green's function and its derivatives, verify that the function*

$$f(x, y, z) = x^3 + 3xy^2 - z - 2,$$

can be recovered on a finite domain from its Laplacian and the values of the function and its derivatives on the boundary.

The grid spacings are chosen to equal those from the gel phantom experiment mentioned in this dissertation, since these Green's filter coefficients have already been calculated. The grid dimensions are chosen to be $256 \times 256 \times 16$, again to match the gel-phantom grid dimensions. The function is evaluated on this discrete grid, as are its Laplacian and first derivatives:

$$\frac{\partial f}{\partial x} = 3x^2 + 3y^2, \qquad \frac{\partial f}{\partial y} = 6xy, \qquad \frac{\partial f}{\partial z} = -1, \qquad \nabla^2 f = 12x.$$

To handle the finite domain, a boxcar window must be introduced. The function should be recoverable with the equation:

$$f = \mathcal{G} * \left(W \nabla^2 f \right) + \underbrace{\mathcal{G} * \left(2 \nabla W \cdot \nabla f + f \nabla^2 W \right)}_{\mathcal{B}}.$$

For the boxcar function, derivatives must be considered in the distributional sense, which introduces δ and δ' distributions. These are what evaluate the function and its first derivatives at the boundary. The boundary term is computed using the expression in Equation (A.2), and the approximated Green's filters. The centre slice of results are shown in Figure A.2. The relative error between the two sides of the equation can be calculated:

$$\text{Error} = \frac{\left\| \mathcal{G} * \left(W \nabla^2 f \right) + \mathcal{B} - W f \right\|}{\| W f \|} \times 100\%$$
$$= 0.019\%.$$

The very small error is a result of the constant term, 2, being correctly determined. The function on this small grid has a relatively large mean compared to its variation. If the constant is removed from the calculation,

$$\text{Error} = \frac{\left\| \mathcal{G} * \left(W \nabla^2 f \right) + \mathcal{B} - W f \right\|}{\| W f + 2 \|} \times 100\%$$
$$= 2.76\%,$$

which is still quite small. Thus, these discrete approximations of $\mathcal{G}$ and its derivatives are able to successfully recover a function from its Laplacian and boundary values.
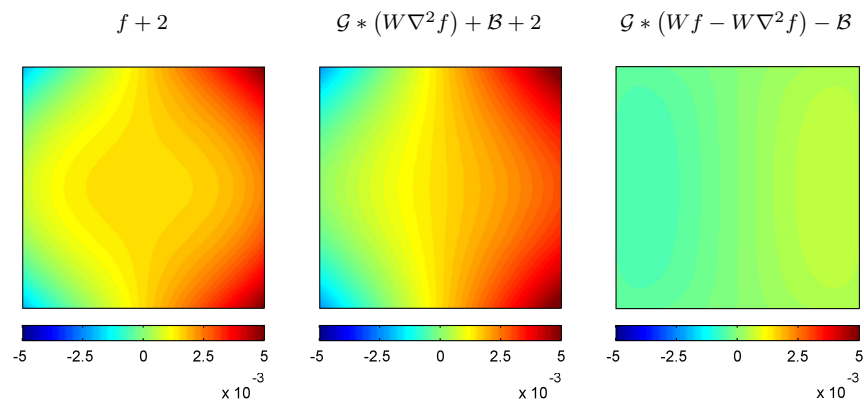
Figure A.2: Test of Green's function approximations. Images depict results from the centre slice, parallel to the xy-plane.

In order to test the accuracy of inversion techniques, a set of simulated data is required. This means the forward problem must be solved. Solving the system of harmonic wave equations in three dimensions requires inverting an incredibly large matrix system. Even for small grids, the inversion is extremely computationally expensive. Instead of tackling the implicit problem, the time-dependent elastic wave equation was simulated:

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \nabla \cdot \left(\mu \left[\nabla \boldsymbol{u} + \nabla^{\mathrm{T}} \boldsymbol{u} \right] \right) + \nabla \left(\lambda \nabla \cdot \boldsymbol{u} \right). \tag{B.1}$$

The explicit scheme requires no matrix inversion, so data can be generated rather quickly, even on large grids. One simply needs to wait until the time-dependent system reaches a quasi-steady state, then the harmonic data can be obtained by transforming the displacement field using an FFT.

The inversion techniques are meant to work on local data, without knowledge of boundary conditions. Thus, the boundary conditions for the forward problem are free. To ensure the system quickly reached a quasi-steady state, perfectly absorbing boundaries were implemented. This was accomplished with the use of an adapted version of Berenger's Perfectly Matched Layer (PML) [4, 5].

B.1 The Perfectly Matched Layer

The idea of the PML is to add an outer layer outside the region of interest in which waves are forced to decay rapidly. As long as this layer does not cause any reflections, all the energy is able to propagate outside the region of interest as if travelling outwards indefinitely. To create this exponential decay, the wave equation and its solutions are analytically extended onto a new complex coordinate system: $\tilde{\boldsymbol{x}} = \boldsymbol{x} + i\boldsymbol{f}(\boldsymbol{x})$. If the original solution can be written as

$$u_i(\boldsymbol{x}, t) = \sum_{\boldsymbol{k}, \omega} e^{i\boldsymbol{k} \cdot \boldsymbol{x} - i\omega t}, \tag{B.2}$$

then the new, extended one becomes

$$u_i(\tilde{\boldsymbol{x}}, t) = \sum_{\boldsymbol{k}, \omega} e^{i\boldsymbol{k} \cdot \boldsymbol{x} - i\omega t + \boldsymbol{k} \cdot \boldsymbol{f}(\boldsymbol{x})}.$$

The complex coordinate function, $\boldsymbol{f}$, allows for the solution to be manipulated. If $\boldsymbol{f} = 0$ in the region of interest (ROI), then the solution is unchanged. Otherwise, waves are forced to decay exponentially as they travel through space. As long as these waves are allowed to decay to zero before they can reflect back, no energy will return to the ROI. The result is a perfectly absorbing boundary condition. The boundary conditions on the outside of the boundary layer are arbitrary, since the waves should have
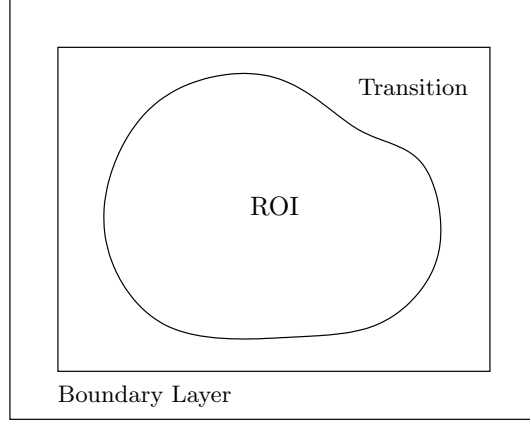
vanished before striking these edges.

Figure B.1: Boundary layer. All parameters should be constant in the boundary layer, so a transition region may be required where they smoothly transition to some fixed values.

To guarantee solutions are of the type described in Equation (B.2), the parameters in the system must be constant. Therefore, a transition region may be required between the ROI and the boundary layer, in which λ, μ and ρ are smoothly transitioned to some constant values. The boundary layer is then added outside the transition region. This setup is depicted in Figure B.1.

The original PML was designed for uncoupled electromagnetic waves. These waves obey the simple relation:

$$\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = \frac{1}{c^2} \nabla^2 \boldsymbol{E},$$

where $\boldsymbol{E}$ is the electric (or magnetic) field, and c is the speed of light. The elastic wave equation is substantially more complex. Not only does have more terms, but the components are coupled and the wave-speed parameters are spatially dependent. Despite these differences, the same concepts can still be applied to derive an absorbing layer.

To implement the boundary layer, the system is split into a system of advection-type equations. For for the electromagnetic wave system discussed by Berenger, this resulted in a system of six equations for each polarization. Since the polarizations are uncoupled, they can be simulated separately. For the full elastic wave equation, the split system consists of the following 31 coupled equations:

$$
\begin{array}{lllll}
\phi_{11,t} = 2\mu\,u_{1,1} & \phi_{12,t} = \mu\,u_{1,2} & \phi_{13,t} = \mu\,u_{1,3} & \phi_{14,t} = \mu\,u_{2,1} & \phi_{15,t} = \mu\,u_{3,1} \\
\phi_{21,t} = \mu\,u_{2,1} & \phi_{22,t} = 2\mu\,u_{2,2} & \phi_{23,t} = \mu\,u_{2,3} & \phi_{24,t} = \mu\,u_{1,2} & \phi_{25,t} = \mu\,u_{3,2} \\
\phi_{31,t} = \mu\,u_{3,1} & \phi_{32,t} = \mu\,u_{3,2} & \phi_{33,t} = 2\mu\,u_{3,3} & \phi_{34,t} = \mu\,u_{3,1} & \phi_{35,t} = \mu\,u_{3,2}
\end{array}
$$

$$
\begin{array}{llll}
\varphi_{11,t} = (\phi_{11} + \psi)_{,1} & \varphi_{12,t} = (\phi_{12} + \phi_{14})_{,2} & \varphi_{13,t} = (\phi_{13} + \phi_{15})_{,3} & \rho\,u_1 = \varphi_{11} + \varphi_{12} + \varphi_{13} \\
\varphi_{21,t} = (\phi_{21} + \phi_{24})_{,1} & \varphi_{22,t} = (\phi_{22} + \psi)_{,2} & \varphi_{23,t} = (\phi_{23} + \phi_{25})_{,3} & \rho\,u_2 = \varphi_{21} + \varphi_{22} + \varphi_{23} \\
\varphi_{31,t} = (\phi_{31} + \phi_{34})_{,1} & \varphi_{32,t} = (\phi_{32} + \phi_{35})_{,2} & \varphi_{33,t} = (\phi_{33} + \psi)_{,3} & \rho\,u_3 = \varphi_{31} + \varphi_{32} + \varphi_{33}
\end{array}
$$

$$\psi_{1,t} = \lambda\, u_{1,1} \qquad\qquad \psi_{2,t} = \lambda\, u_{2,2} \qquad\qquad \psi_{3,t} = \lambda\, u_{3,3} \qquad\qquad \psi = \psi_1 + \psi_2 + \psi_3$$

This system is analytically extended onto the complex coordinates, then transformed back to the original real-valued ones. To see the effect of this transformation, consider one advection-type equation:

$$\frac{\partial v}{\partial t} = \frac{\partial w}{\partial \tilde{x}}$$

If v is assumed time-harmonic, this can be expressed in the original real-valued coordinates:

$$\frac{\partial v}{\partial t} = \frac{\partial w}{\partial x}\left(\frac{1}{1 + i\frac{\partial f_i}{\partial x_i}}\right)$$

$$i\omega v\left(1 + i\frac{\partial f_i}{\partial x_i}\right) = \frac{\partial w}{\partial x_i}$$

$$\frac{\partial v}{\partial t} - \omega\frac{\partial f_i}{\partial x_i}v = \frac{\partial w}{\partial x_i}$$

$$\frac{\partial v}{\partial t} + \eta_i v = \frac{\partial w}{\partial x_i},$$

The function η_i is responsible for exponential decay of the waves travelling in the ith direction in the absorbing layer. These functions should be positive in the boundary layer to force exponential decay, and zero in the region of interest so as to not affect the original solution. With this incorporation of the η functions, a perfectly absorbing boundary condition is enforced, allowing the system to quickly reach a quasi-steady state. Each advection-type equation is transformed in this manner, resulting in a new set of 31 equations. This system is then simulated with an explicit scheme until the displacement field reaches a quasi-steady state.

B.2   Numerics

The system of advection-type equations is marched forward using Euler time-stepping. The coupling is handled by staggering the equations by half a time-step using the following procedure:

1. Step all ϕ's and ψ's forward by half a time-step.

2. Step all φ's forward by a full time-step.

3. Update the displacement field, $\boldsymbol{u}$, and apply boundary/forcing conditions.

4. Step all ϕ's and ψ's forward by another half time-step.

At each half step, a trapezoidal rule is employed, replacing $v_{,t}^{(n)}$ with $\frac{1}{2}v_{,t}^{(n+1/2)} + \frac{1}{2}v_{,t}^{(n-1/2)}$. For stability, the explicit scheme requires the Courant-Friedrichs-Lewy (CFL) condition to hold:

$$\Delta t < \frac{\min\{\Delta x, \Delta y, \Delta z\}}{\max\{1, 2\mu, \lambda\}}.$$

The decay functions in the boundary layer are chosen to have the following form:

$$\eta_i(x) = \begin{cases} \alpha(x_l - x)^3, & x < x_l \\ 0, & x_l \leq x \leq x_r \\ \alpha(x - x_r)^3, & x > x_r \end{cases}$$

where $[x_l, x_r]$ is the domain in the ith spatial dimension, and α scales the rate of exponential decay. With these, the forward scheme with the absorbing layer is iterated as follows:

$$\phi_{11}^{(n+1/2)} = 2\mu \frac{2\Delta t}{4 + \eta_1 \Delta t} \left(u_1^{(n)}\right)_{,1} + \frac{4 - \eta_1 \Delta t}{4 + \eta_1 \Delta t} \phi_{11}^{(n)}$$

$$\vdots$$

$$\psi_1^{(n+1/2)} = \lambda \frac{2\Delta t}{4 + \eta_1 \Delta t} \left(u_1^{(n)}\right)_{,1} + \frac{4 - \eta_1 \Delta t}{4 + \eta_1 \Delta t} \psi_1^{(n)}$$

$$\vdots$$

$$\psi^{(n+1/2)} = \psi_1^{(n+1/2)} + \psi_2^{(n+1/2)} + \psi_3^{(n+1/2)}$$

$$\varphi_{11}^{(n+1)} = \frac{1}{\rho} \frac{2\Delta t}{2 + \eta_1 \Delta t} \left(\phi_{11}^{(n+1/2)} + \psi^{(n+1/2)}\right)_{,1} + \frac{2 - \eta_1 \Delta t}{2 + \eta_1 \Delta t} \varphi_{11}^{(n)}$$

$$\vdots$$

$$u_1^{(n+1)} = \varphi_{11}^{(n+1)} + \varphi_{12}^{(n+1)} + \varphi_{13}^{(n+1)}$$
$$u_2^{(n+1)} = \varphi_{21}^{(n+1)} + \varphi_{22}^{(n+1)} + \varphi_{23}^{(n+1)}$$
$$u_3^{(n+1)} = \varphi_{31}^{(n+1)} + \varphi_{32}^{(n+1)} + \varphi_{33}^{(n+1)}$$

$\langle$ Apply boundary conditions and forcing $\rangle$

$$\phi_{11}^{(n+1)} = 2\mu \frac{2\Delta t}{4 + \eta_1 \Delta t} \left(u_1^{(n+1)}\right)_{,1} + \frac{4 - \eta_1 \Delta t}{4 + \eta_1 \Delta t} \phi_{11}^{(n+1/2)}$$

$$\vdots$$

$$\psi_1^{(n+1)} = \lambda \frac{2\Delta t}{4 + \eta_1 \Delta t} \left(u_1^{(n+1)}\right)_{,1} + \frac{4 - \eta_1 \Delta t}{4 + \eta_1 \Delta t} \psi_1^{(n+1/2)}$$

$$\vdots$$

$$\psi^{(n+1)} = \psi_1^{(n+1)} + \psi_2^{(n+1)} + \psi_3^{(n+1)}.$$

For the spatial derivatives, fourth-order finite differencing was used. Centred differencing led to checkerboarding artifacts, since the approximated $f'(\boldsymbol{x}_o)$ does not depend on $f(\boldsymbol{x}_o)$ in an even-order centred finite difference scheme. This effect is shown in Figure B.2. To avoid this, the stencils are shifted over

by one point. Spatial derivatives for the ϕ's and ψ's were estimated using a shifted-forward stencil, and derivatives of the φ's were estimated using a shifted-backward stencil. With this shifted scheme, the checkerboard artifacts are shown to be removed.
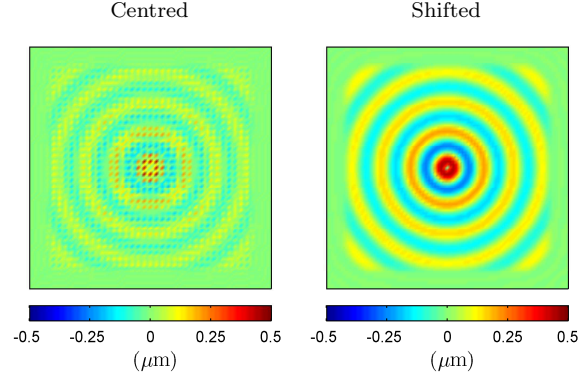

Figure B.2: Checker-boarding caused by centred differencing is removed by the shifted-forward/backward scheme.

For the simulated displacement fields in this dissertation, the following parameters are used:

· Grid: $128 \times 128 \times 24$, plus an additional 12 points on each side for the boundary layer.
· Grid Spacings: $\Delta x = 1/1280$, $\Delta y = 1/1280$, $\Delta z = 1/500$.
· Time Step: $\Delta t = 1/20000$, satisfying the CFL condition for $\mu_{\max} = 7$.
· Boundary Layer: $\alpha = 10000$, Dirichlet zero boundary conditions on outside.
· Forcing: $u_1 = 0$, $u_2 = 0$, $u_3 = \sin(2\pi 100 t)$, applied to $x = 0$, just inside the boundary layer.
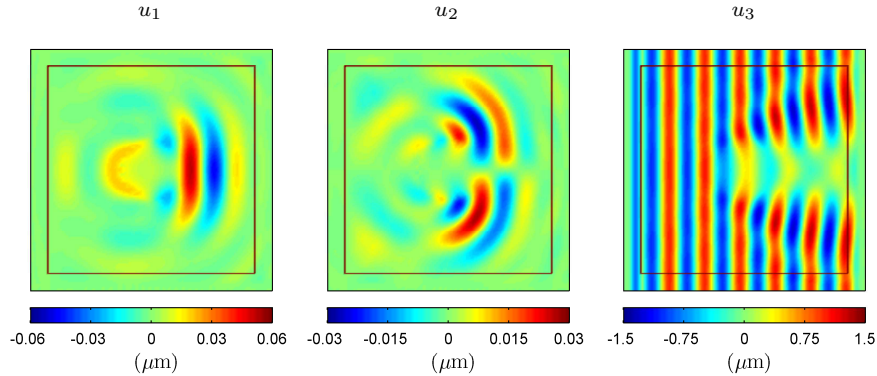
Figure B.3: Forward simulation, spherical inclusion. The centre slice of the displacement field, parallel to the xy-plane, is shown. The edge of the boundary layer is marked.

A slice of the displacement field for the spherical inclusion simulation is shown in Figure B.3. Notice that only waves propagating outward in the PML are forced to decay; waves travelling parallel to the boundary are unaffected. Non-zero forcing is only applied to displacements in the z-polarization, meaning

displacements in the other two orientations are caused by the coupling terms in the equations of motion. The boundary layer is removed before the data is used in the inversion procedures.

PDE Systems and Uniqueness

One of the concerns in magnetic resonance elastography is the uniqueness of the solution. Theoretically, there is a 'true' distribution of the stiffness in a soft tissue. To guarantee that this can be found, there must be enough information available to completely determine the unique solution. Due to the non-invasive nature of the experiment, no stiffness values can be directly measured. The only measurable quantity is the displacement field. This restricts the type of available information.

In the tissue model, the shear modulus is related to the displacement field through a system of partial differential equations. A single differential equation, alone, has an infinite number of solutions. The collection of these are referred to as its family of solutions, which describes a solution surface. To single a particular element out, extra conditions need to be imposed. Typically, this is done by specifying initial or boundary conditions. Only a subset of the family will satisfy the conditions, narrowing the set of admissible solutions. When enough restrictions are imposed, the set can be reduced to a single element, a unique solution. Unfortunately, since the shear modulus cannot be directly measured, boundary conditions are unknown in the MRE experiment. One may guess at conditions, but then there is no guarantee they lead to the correct solution.

Applying boundary conditions is not the only way to narrow the set of solutions. Instead, a regularizing condition can be imposed, such as selecting the family member with the minimum variation. Of all the functions in the solution family, there will only be a subset that have this minimum variation. Again, with enough conditions, the number of admissible solutions can be reduced to one.

Most direct methods use a makeshift form of regularization. By assuming local homogeneity, one searches for an approximate solution that varies the least. The desired function is assumed locally constant, so is pulled out of all derivative terms and isolated:

$$h = L(fg) \approx fL(g) \qquad \implies \qquad f \approx h/L(g),$$

where f is the desired quantity, L is a linear differential operator, and g and h are known functions. Whenever f does not have zero derivatives, the simplified equation is invalid. The resulting estimate no longer satisfies the original differential relationship. In MRE, the shear modulus is expected to change. In fact, this is the purpose of the procedure; a change in modulus indicates a change in stiffness, which is the indicator for pathological tissue. A better approach is to use the original differential equations, but enforce the regularization condition through optimization techniques. In this way, a regularized solution can be found that still satisfies the original system. This is the basis of the IQO methods. Still, the solution is only as accurate as the validity of the regularizing assumptions.

There is a third way the set of solutions can be limited. This method is directly applicable to magnetic resonance elastography, and uses the fact that there are more PDEs than unknown functions in the relationship between the shear modulus and the displacement field.

C.1   PDE Systems

Consider a system of consistent differential equations:

$$L_1(f) = h_1,$$
$$L_2(f) = h_2,$$
$$\vdots$$
$$L_N(f) = h_N,$$

where f is the desired function, L_i is a linear differential operator, and h_i is a known function. Without prescribing some form of conditions, each differential equation has its own family of solutions, $\mathcal{S}_i$. The desired function, f, is common to all equations, so therefore must be a member of each family. That is,

$$f \in \mathcal{S}_1 \cap \mathcal{S}_2 \cap \cdots \cap \mathcal{S}_N.$$

This intersection can be quite small. If further consistent, linearly independent equations can be added to the system, then the set of admissible solutions can be narrowed even further. In some cases, this can be repeated until there is a unique solution. In this way, a system of differential equations can have a unique solution without the need for any initial, boundary, or regularizing conditions.

The idea is easily generalized to finding multiple functions, f_1 and f_2. For two functions, a pair of linear differential equations is required in order to have a family of solutions, $(f_1, f_2) \in (\mathcal{S}_{11}, \mathcal{S}_{21})$. When further equations are added, further pairs are made, each having their own solution family. For a system with N pairs of equations, the two desired functions must satisfy:

$$(f_1, f_2) \in \bigcap_{i=1}^{N}(\mathcal{S}_{1i}, \mathcal{S}_{2i}).$$

With enough linearly independent equations, and under certain conditions, this solution pair can be unique.

The full conditions under which a unique solution may exist has yet to be explored by the author. One condition is that at least one of the equations must contain a term of the form $c(x)f(x)$ (i.e. there must be at least one term involving f not acted on by a derivative operator). If only the derivatives of f appear in the system of equations, then the solution can at best be unique up to the addition of a constant (since the system is unchanged when f is modified by a constant).

This type of theory lies in the field of *differential topology*. Further investigation of this area of mathematics is required in order to develop a full set of conditions under which the solution may be unique for the MRE problem. A potentially useful theorem is that of Frobenius, applied to a homogeneous system of first-order partial differential equations. A full development of this type of theory, applied to mechanics, can by found in Abraham and Marsden [1].

Theorem C.1 (Frobenius Theorem, simplified)**.** *Let $f_{ik}(\boldsymbol{x})$ be a collection of $C^1(\mathbb{R}^n)$ functions such*

that the $r \times n$ matrix $[f_{ik}]$ has rank r. The system of homogeneous differential equations

$$L_1 u = \sum_{k=1}^{n} f_{1k} \frac{\partial u}{\partial x_k} = 0,$$

$$L_2 u = \sum_{k=1}^{n} f_{2k} \frac{\partial u}{\partial x_k} = 0,$$

$$\vdots$$

$$L_r u = \sum_{k=1}^{n} f_{rk} \frac{\partial u}{\partial x_k} = 0,$$

has a collection of $n - r$ solutions with linearly independent gradients if and only if the following involutivity condition holds:

$$(L_i L_j - L_j L_i) u(\boldsymbol{x}) = \sum_{k=1}^{r} c_{ijk}(\boldsymbol{x}) L_k u(\boldsymbol{x}),$$

for every $i, j \in \{1, 2, \ldots, r\}$ and all $u \in C^2(\mathbb{R}^n)$.

A result of Frobenius' theorem is that if there are n linearly independent equations that satisfy the involutivity condition, then there are zero solutions that have linearly independent gradients, meaning the solution must be the constant function. This result should be clear, since for n linearly independent equations, each of $\frac{\partial u}{\partial x_i}$ can be isolated using Gaussian elimination techniques, and will be equal to zero. The only functions that have uniformly zero gradients are constant functions. In this instance, a set of linearly independent differential equations have an intersecting family of solutions containing only functions of the form $u = c \in \mathbb{C}^n$.

C.1.1   Application to MRE

In order to apply this concept of intersecting families (or solution surfaces), the PDE system must be overdetermined (in the sense that there must be more equations than unknown functions). This is true for MRE. The expanded viscoelastic equation of motion for an isotropic continuum is the following:

$$-\omega^2 \mathbf{U} = \mathcal{M} \nabla \cdot \left[\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right] + \nabla \mathcal{M} \cdot \left[\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right] + \Lambda \nabla \nabla \cdot \mathbf{U} + \nabla \Lambda \left(\nabla \cdot \mathbf{U} \right),$$

which can be re-written in a matrix form:

$$-\omega^2 \mathbf{U} = \underbrace{\left[\nabla \cdot \left[\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right] \quad \left[\nabla \mathbf{U} + \nabla^{\mathrm{T}} \mathbf{U} \right] \quad \nabla \nabla \cdot \mathbf{U} \quad \begin{bmatrix} \nabla \cdot \mathbf{U} \\ \nabla \cdot \mathbf{U} \\ \nabla \cdot \mathbf{U} \end{bmatrix} \right]}_{=A(\boldsymbol{x})} \begin{bmatrix} \mathcal{M} \\ \nabla \mathcal{M} \\ \Lambda \\ \nabla \Lambda \end{bmatrix}. \qquad \text{(C.1)}$$

This is a system of three differential equations, each with a common $\mathcal{M}$ and Λ. Note that the equations do contain terms of the form $c(\boldsymbol{x})\mathcal{M}(\boldsymbol{x})$ and $d(\boldsymbol{x})\Lambda(\boldsymbol{x})$. This allows for the possibility of determining a unique $\mathcal{M}$ and Λ, without making assumptions. The rows of A are not expected to be linearly dependent, but they are expected to produce a consistent differential system (otherwise there is a major flaw in the model). There are three pairs of equations for the two unknown functions, so there are three paired families of solutions. These intersect in some smaller set, defining the set of admissible solutions.

Note that the coefficients of the differential operator in Equation (C.1), $A(\boldsymbol{x})$, are spatially varying, and are dependent on the displacement field. If a second displacement field, $\mathbf{U}_2$, is measured, it will create its own differential operator with coefficients $A_2(\boldsymbol{x})$. As long as the domain of the new field is spatially aligned with the original, the two unknown parameters are common to both systems. The systems can therefore be merged, potentially contracting the set admissible solutions. This can be repeated until the solution is unique. Linearly independent displacement fields are easily obtained by changing the location of the mechanical actuator in the experiment.

The basis of the PIQO method derived in Chapter 6 is that Equation (C.1) can be solved algebraically, as long as there are enough equations. If the Λ term is ignored, then only four consistent, linearly independent differential equations are required. Otherwise, eight are needed. With these, $\mathcal{M}$ and Λ are fully determined, even though no external conditions have been applied. This is an example of when solution families intersect at a single element.

The PIQO method ignores the relationships between $\mathcal{M}$ and $\nabla\mathcal{M}$. If these relations are considered, as done in the BIQO method, perhaps even fewer equations are required. In fact, through trials, it was found that only two equations were enough to solve for both $\mathcal{M}$ and $\nabla\mathcal{M}$, without needing to add extra conditions. This suggests that measuring a single displacement field is sufficient to solve for $\mathcal{M}$, without having to make any regularizing assumptions.

If the longitudinal component is ignored, the differential equations for $\mathcal{M}$ are of the form:

$$\nabla\!\cdot\!(\mathcal{M}\boldsymbol{g}) = h,$$

where $\boldsymbol{g}$ is a known row vector field, and h is a known function. The following propositions and theorems explore when a solution to a system of such equations might be unique, without requiring any initial, boundary or regularizing conditions.

For the remainder of this chapter, consider $f : \mathbb{R}^n \to \mathbb{C}$, $\boldsymbol{g}_i : \mathbb{R}^n \to \mathbb{C}^n$, and $h_i : \mathbb{R}^n \to \mathbb{C}$, where each h_i is continuous, and f and $\boldsymbol{g}_i$ are continuously differentiable.

Proposition C.1. *There exists two fields, $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, such that the following system of homogeneous partial differential equations has a unique solution:*

$$\nabla\!\cdot\!(f\boldsymbol{g}_1) = 0,$$
$$\nabla\!\cdot\!(f\boldsymbol{g}_2) = 0,$$

without the need to prescribe external conditions.

Proof. First of all, note that the system is consistent; the zero function satisfies both equations. If the solution is unique, then it must be $f = 0$.

Let $\boldsymbol{g}_1(\boldsymbol{x}) = (g_1(x_1), 0, \ldots, 0)$, and $\boldsymbol{g}_2(\boldsymbol{x}) = (g_2(x_1), 0, \ldots, 0)$ be two non-zero vector fields, with $g_2 \neq c\, g_1$ for any $c \in \mathbb{C}$. Inserting these into the differential equation results in:

$$\frac{\partial}{\partial x_1}\left(f(\boldsymbol{x})g_1(x_1)\right) = 0, \qquad\qquad \frac{\partial}{\partial x_1}\left(f(\boldsymbol{x})g_2(x_1)\right) = 0.$$

The solutions are easily found to be

$$f(\boldsymbol{x}) = \frac{h_1(x_2, x_3, \ldots, x_n)}{g_1(x_1)} \qquad \text{and} \qquad f(\boldsymbol{x}) = \frac{h_2(x_2, x_3, \ldots, x_n)}{g_2(x_1)},$$

where h_1 and h_2 are arbitrary functions not dependent on x_1. The solution, f, must belong to both families. At the intersection,

$$\frac{h_1(x_2, x_3, \ldots, x_n)}{g_1(x_1)} = \frac{h_2(x_2, x_3, \ldots, x_n)}{g_2(x_1)}.$$

Assume h_1 is not the zero function. Then,

$$\frac{g_2(x_1)}{g_1(x_1)} = \frac{h_2(x_2, x_3, \ldots, x_n)}{h_1(x_2, x_3, \ldots, x_n)}.$$

The left-hand side of this expression is only a function of x_1, but the right-hand side is *not* a function of x_1. This can only hold if both sides are constant. Therefore,

$$\frac{g_2(x_1)}{g_1(x_1)} = c \qquad \Longrightarrow \qquad g_2 = cg_1.$$

But g_1 and g_2 are linearly independent by construction. This is a contradiction. Therefore,

$$h_1(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^3$$
$$\Longrightarrow \quad f(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^3.$$

$\square$

A unique solution, $f = 0$, was determined for the two constructed fields, $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, without imposing external conditions. In this case, the two separate families of solutions were found, and the intersection was determined. This idea is difficult to apply to non-homogeneous problems, since one must first find a particular solution. Instead, an algebraic proof is presented.

Proposition C.2. *There exists two fields, $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, and two functions, h_1 and h_2, such that the following non-homogeneous system of partial differential equations has a unique solution:*

$$\nabla\cdot(f\boldsymbol{g}_1) = h_1,$$
$$\nabla\cdot(f\boldsymbol{g}_2) = h_2,$$

without the need to prescribe external conditions.

Proof. Let $\boldsymbol{g}_1(\boldsymbol{x}) = (g_1(x_1), 0, \ldots, 0)$, and $\boldsymbol{g}_2(\boldsymbol{x}) = (g_2(x_1), 0, \ldots, 0)$ be two non-zero vector fields, with $g_2 \neq c\, g_1$ for any $c \in \mathbb{C}$. The linear independence of g_1 and g_2 means that:

$$\frac{\partial}{\partial x_1}\left(\frac{g_2}{g_1}\right) \neq 0$$

$$\implies \quad \frac{\partial g_1}{\partial x_1} g_2 - g_1 \frac{\partial g_2}{\partial x_1} \neq 0.$$

Here, zero refers to the zero function. The expression can be zero at certain points $\boldsymbol{x} \in \mathbb{R}^3$, but not at all points. Expanding the system of differential equations and converting to matrix form,

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & g_1 \\ \frac{\partial g_2}{\partial x_1} & g_2 \end{bmatrix} \begin{bmatrix} f \\ \frac{\partial f}{\partial x_1} \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

$$\implies \quad \begin{bmatrix} f \\ \frac{\partial f}{\partial x_1} \end{bmatrix} = \frac{1}{\frac{\partial g_1}{\partial x_1} g_2 - g_1 \frac{\partial g_2}{\partial x_1}} \begin{bmatrix} g_2 h_1 - g_1 h_2 \\ -\frac{\partial g_2}{\partial x_1} h_1 + \frac{\partial g_1}{\partial x_1} h_2 \end{bmatrix}.$$

The denominator was shown to not be the zero function, so this expression is defined for some values of $\boldsymbol{x} \in \mathbb{R}^3$. If the system is consistent, then the unique solution must be

$$f = \frac{g_2 h_1 - g_1 h_2}{\frac{\partial g_1}{\partial x_1} g_2 - g_1 \frac{\partial g_2}{\partial x_1}}.$$

For consistency, the functions h_1 and h_2 must be chosen to satisfy

$$\frac{\partial}{\partial x}\left(\frac{g_2 h_1 - g_1 h_2}{\frac{\partial g_1}{\partial x_1} g_2 - g_1 \frac{\partial g_2}{\partial x_1}}\right) = \frac{-\frac{\partial g_2}{\partial x_1} h_1 + \frac{\partial g_1}{\partial x_1} h_2}{\frac{\partial g_1}{\partial x_1} g_2 - g_1 \frac{\partial g_2}{\partial x_1}}$$

As a specific example, take $g_1 = \sin(x_1)$, $g_2 = \cos(x_1)$, $h_1 = -\cos(x_1)$, $h_2 = \sin(x_1)$. It is easily shown that the expression for f exists and is continuous everywhere, and that the system is consistent. Therefore, it is possible for two non-homogeneous partial differential equations to have a unique solution, without needing to prescribe boundary conditions. $\qquad \square$

Theorem C.2. *For any $n+1$ vector fields, $\boldsymbol{g}_i : \mathbb{R}^n \to \mathbb{C}^n$, that satisfy:*

$$\begin{vmatrix} \nabla \cdot \boldsymbol{g}_1 & \boldsymbol{g}_1 \\ \nabla \cdot \boldsymbol{g}_2 & \boldsymbol{g}_2 \\ & \vdots \\ \nabla \cdot \boldsymbol{g}_{n+1} & \boldsymbol{g}_{n+1} \end{vmatrix} \neq 0, \tag{C.2}$$

there exists a region of non-zero measure, Γ, over which the consistent system of partial differential

equations

$$\nabla \cdot (f\boldsymbol{g}_1) = h_1,$$
$$\nabla \cdot (f\boldsymbol{g}_1) = h_2,$$
$$\vdots$$
$$\nabla \cdot (f\boldsymbol{g}_{n+1}) = h_{n+1},$$

has a unique solution, without requiring external conditions.

Proof. The proof is straightforward. Convert the system of partial differential equations into a matrix system:

$$\begin{bmatrix} \nabla \cdot \boldsymbol{g}_1 & \boldsymbol{g}_1 \\ \nabla \cdot \boldsymbol{g}_2 & \boldsymbol{g}_2 \\ & \vdots \\ \nabla \cdot \boldsymbol{g}_{n+1} & \boldsymbol{g}_{n+1} \end{bmatrix} \begin{bmatrix} f \\ \nabla f \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{n+1} \end{bmatrix}.$$

By conditions of the theorem, the matrix of coefficients is invertible for some values of $\boldsymbol{x} \in \mathbb{R}^3$. Due to the continuity of each $\boldsymbol{g}_i$, the matrix is also invertible in some neighbourhood of these points. The union of these neighbourhoods make up Γ. In Γ, the solution can be found by taking the first entry of

$$\begin{bmatrix} f \\ \nabla f \end{bmatrix} = \begin{bmatrix} \nabla \cdot \boldsymbol{g}_1 & \boldsymbol{g}_1 \\ \nabla \cdot \boldsymbol{g}_2 & \boldsymbol{g}_2 \\ & \vdots \\ \nabla \cdot \boldsymbol{g}_{n+1} & \boldsymbol{g}_{n+1} \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{n+1} \end{bmatrix}.$$

If the original system is consistent, then this transformed system is also consistent. Thus, the solution is uniquely determined within some domain, Γ, which has a non-zero measure. $\square$

Theorem C.3. *Consider a consistent system of the form:*

$$\nabla \cdot (f\boldsymbol{g}_1) = h_1,$$
$$\nabla \cdot (f\boldsymbol{g}_2) = h_2,$$
$$\vdots$$
$$\nabla \cdot (f\boldsymbol{g}_N) = h_N.$$

For any point, $\boldsymbol{x}_o$, such that

$$\mathrm{rank}\left\{\begin{bmatrix}\boldsymbol{g}_1(\boldsymbol{x}_o)\\\boldsymbol{g}_2(\boldsymbol{x}_o)\\\vdots\\\boldsymbol{g}_N(\boldsymbol{x}_o)\end{bmatrix}\right\}\quad<\quad\mathrm{rank}\left\{\begin{bmatrix}\nabla\cdot\boldsymbol{g}_1(\boldsymbol{x}_o)&\boldsymbol{g}_1(\boldsymbol{x}_o)\\\nabla\cdot\boldsymbol{g}_2(\boldsymbol{x}_o)&\boldsymbol{g}_2(\boldsymbol{x}_o)\\&\vdots\\\nabla\cdot\boldsymbol{g}_N(\boldsymbol{x}_o)&\boldsymbol{g}_N(\boldsymbol{x}_o)\end{bmatrix}\right\},$$

the solution, f, is unique.

Proof. At each point $\boldsymbol{x}_o$, a local Gaussian elimination technique can be used to isolate $f(\boldsymbol{x}_o)$ in the equations. In particular, there is a set of row operations, which can be represented by multiplication with an $N \times N$ matrix $R(\boldsymbol{x}_o)$, such that

$$R\begin{bmatrix}\nabla\cdot\boldsymbol{g}_1(\boldsymbol{x}_o)&\boldsymbol{g}_1(\boldsymbol{x}_o)\\\nabla\cdot\boldsymbol{g}_2(\boldsymbol{x}_o)&\boldsymbol{g}_2(\boldsymbol{x}_o)\\&\vdots\\\nabla\cdot\boldsymbol{g}_N(\boldsymbol{x}_o)&\boldsymbol{g}_N(\boldsymbol{x}_o)\end{bmatrix}\begin{bmatrix}f(\boldsymbol{x}_o)\\\nabla f(\boldsymbol{x}_o)\end{bmatrix}=\begin{bmatrix}f(\boldsymbol{x}_o)\\\vdots\end{bmatrix}=R\begin{bmatrix}h_1(\boldsymbol{x}_o)\\h_2(\boldsymbol{x}_o)\\\vdots\\h_N(\boldsymbol{x}_o)\end{bmatrix}.$$

The rank condition allows $f(\boldsymbol{x}_o)$ to be isolated in one of the equations, implying it has a unique value. $\qquad\square$

Corollary C.3.1. *When the longitudinal component is ignored, the complex shear modulus can be uniquely determined at all points satisfying:*

$$\mathrm{rank}\left\{\left[\nabla\mathbf{U}+\nabla^T\mathbf{U}\right]\right\}<\mathrm{rank}\left\{\begin{bmatrix}\nabla\cdot\left[\nabla\mathbf{U}+\nabla^T\mathbf{U}\right]&\left[\nabla\mathbf{U}+\nabla^T\mathbf{U}\right]\end{bmatrix}\right\}.$$

Proof. The result follows by applying Theorem C.3 to the system in Equation (C.1), setting Λ and $\nabla\Lambda$ equal to zero. $\qquad\square$

Theorem C.4. *Consider the homogeneous system:*

$$\begin{aligned}\nabla\cdot(f\boldsymbol{g}_1)&=0,\\\nabla\cdot(f\boldsymbol{g}_2)&=0,\\&\vdots\\\nabla\cdot(f\boldsymbol{g}_n)&=0.\end{aligned}\tag{C.3}$$

In any open region, Γ, where $[\boldsymbol{g}_i]$ has full rank and the vector field

$$F=\begin{bmatrix}\boldsymbol{g}_1\\\boldsymbol{g}_2\\\vdots\\\boldsymbol{g}_n\end{bmatrix}^{-1}\begin{bmatrix}\nabla\cdot\boldsymbol{g}_1\\\nabla\cdot\boldsymbol{g}_2\\\vdots\\\nabla\cdot\boldsymbol{g}_n\end{bmatrix}$$

is not *conservative, then the solution to the homogeneous system in Equation* (C.3) *is unique:* $f(\boldsymbol{x}) = 0,\ \forall \boldsymbol{x} \in \Gamma$.

Proof. Take $\boldsymbol{x}_o \in \Gamma$. Assume the solution is not unique at this point. If the solution is not unique, then there exists a function, $\tilde{f}$, that satisfies the homogeneous system, with $\tilde{f}(\boldsymbol{x}_o) \neq 0$. Due to the continuity of $\tilde{f}$, there is a neighbourhood, $B_r(\boldsymbol{x}_o) \subset \Gamma$, over which

$$\tilde{f}(\boldsymbol{x}) \neq 0, \quad \forall \boldsymbol{x} \in B_r(\boldsymbol{x}_o).$$

In this neighbourhood, the solution can be expressed in terms of an exponential:

$$\tilde{f}(\boldsymbol{x}) = \mathrm{e}^{v(\boldsymbol{x})}, \tag{C.4}$$

where v is a continuous function. The existence of such a v is shown by letting $v = \log(\tilde{f})$. For complex-valued functions, one must evaluate the complex logarithm in such a way that the output is continuous. This can be achieved by gluing together appropriate branches of the complex logarithm. Since $\tilde{f}$ is not zero over $B_r(\boldsymbol{x}_o)$, v is well-defined. Substituting (C.4) into (C.3), the system is transformed:

$$
\begin{aligned}
(\nabla{\cdot}\boldsymbol{g}_1 + \nabla v \cdot \boldsymbol{g}_1)\mathrm{e}^v &= 0 \\
(\nabla{\cdot}\boldsymbol{g}_2 + \nabla v \cdot \boldsymbol{g}_2)\mathrm{e}^v &= 0 \\
&\vdots \\
(\nabla{\cdot}\boldsymbol{g}_n + \nabla v \cdot \boldsymbol{g}_n)\mathrm{e}^v &= 0
\end{aligned}
\qquad \Longrightarrow \qquad
\nabla v = -
\begin{bmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_n \end{bmatrix}^{-1}
\begin{bmatrix} \nabla{\cdot}\boldsymbol{g}_1 \\ \nabla{\cdot}\boldsymbol{g}_2 \\ \vdots \\ \nabla{\cdot}\boldsymbol{g}_n \end{bmatrix} = -F.
$$

To solve this new system, one is searching to find a scalar potential function, v, such that $\nabla v = -F$. Since F is *not* a conservative field over Γ, no such scalar potential function exists. Therefore, it must be that $f(\boldsymbol{x}_o) = 0$. $\qquad\square$

Corollary C.4.1. *Consider a consistent system of the form:*

$$
\begin{aligned}
\nabla{\cdot}(f\boldsymbol{g}_1) &= h_1, \\
\nabla{\cdot}(f\boldsymbol{g}_2) &= h_2, \\
&\vdots \\
\nabla{\cdot}(f\boldsymbol{g}_n) &= h_n.
\end{aligned}
\tag{C.5}
$$

In any open region, Γ, where $[\boldsymbol{g}_i]$ has full rank and the vector field

$$
F =
\begin{bmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_n \end{bmatrix}^{-1}
\begin{bmatrix} \nabla{\cdot}\boldsymbol{g}_1 \\ \nabla{\cdot}\boldsymbol{g}_2 \\ \vdots \\ \nabla{\cdot}\boldsymbol{g}_n \end{bmatrix}
$$

is not *conservative, then the solution to the system in* (C.5) *is unique.*

Proof. If the system of differential equations is consistent, then there exists a particular solution, f_p.

This solution is unique up to the addition of a solution to the homogeneous problem. According to Theorem C.4, in any regions where F is not conservative, the homogeneous problem has the unique solution: $f_h(\boldsymbol{x}) = 0$, $\forall \boldsymbol{x} \in \Gamma$. Thus, in these regions, the inhomogeneous problem must also have a unique solution: $f(\boldsymbol{x}) = f_p(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \Gamma$. $\qquad\square$

Corollary C.4.2. *Consider the viscoelastic wave equation with no longitudinal component:*

$$-\omega^2 \, \mathbf{U} = \begin{bmatrix} \nabla \cdot [\nabla \mathbf{U} + \nabla^T \mathbf{U}] & [\nabla \mathbf{U} + \nabla^T \mathbf{U}] \end{bmatrix} \begin{bmatrix} \mathcal{M} \\ \nabla \mathcal{M} \end{bmatrix}. \tag{C.6}$$

In any open region, Γ, where $[\nabla \mathbf{U} + \nabla^T \mathbf{U}]$ is invertible and

$$\nabla \times [\nabla \mathbf{U} + \nabla^T \mathbf{U}]^{-1} \, \nabla \cdot [\nabla \mathbf{U} + \nabla^T \mathbf{U}] \neq 0,$$

the solution for the complex shear modulus is unique.

Proof. In three spatial dimensions, if $\nabla \times F \neq 0$, then F is not conservative. The result then follows from Corollary C.4.1. $\qquad\square$

For the shear modulus estimation problem, Proposition C.2 suggests that two linearly independent equations can sometimes be sufficient to completely determine $\mathcal{M}$, and Theorem C.2 suggests that at most four linearly independent equations are required. Corollaries C.3.1 and C.4.2 give two conditions for which the solution for the shear modulus is unique if the longitudinal component of the waves is negligible. For further conditions, or to consider uniqueness for the full system in Equation (C.1), further investigation of the theory of differential topology is required.

[1] R. Abraham and J. E. Marsden. *Foundations of Mechanics*. Addison–Wesley, California, 2nd edition, 1987. 167

[2] L. Axel and L. Dougherty. MR imaging of motion with spatial modulation of magnetization. *Radiology*, 171:841–845, Jun 1989.

[3] P. E. Barbone and J. C. Bamber. Quantitative elasticity imaging: what can and cannot be inferred from strain images. *Physics in Medicine and Biology*, 47(12):2147–2164, 2002. 5

[4] J.-P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200, Mar 1994. 160

[5] J.-P. Berenger. *Perfectly Matched Layer (PML) for Computational Electromagnetics*. Morgan & Claypool Publishers, 2007. 160

[6] J. Bishop and D. Plewes. An alternate method for calculating elastic properties of breast tissue. In *Proc. Sixth Meeting of the ISMRM*, Sydney, 1998. 8

[7] J. Bishop, G. Poole, M. Leitch, and D. B. Plewes. Magnetic resonance imaging of shear wave propagation in excised tissue. *J Magn Reson Imaging*, 8:1257–1265, 1998. 10

[8] J. Bishop, A. Samani, J. Sciarretta, and D. B. Plewes. Two-dimensional MR elastography with linear inversion reconstruction: methodology and noise analysis. *Phys Med Biol*, 45:2081–2091, Aug 2000.

[9] G. Box. Robustness in the strategy of scientific model building. In R. Launer and G. Wilkinson, editors, *Robustness in Statistics*, New York, 1979. Academic Press. 44

[10] W. Bynum and R. Porter. *Medicine and the Five Senses*. Cambridge University Press, 2005. 1

[11] M. M. Doyley, P. M. Meaney, and J. C. Bamber. Evaluation of an iterative reconstruction method for quantitative elastography. *Phys Med Biol*, 45:1521–1540, Jun 2000. 8

[12] C. S. Drapaca. Direct inversion using potentials. unpublished, 2006. 26

[13] R. L. Ehman. Mayo Clinic Research: Magnetic Resonance Elastography, Jan 2007. `http://mayoresearch.mayo.edu/ehman_lab/mre.cfm`. 6, 74, 119

[14] M. E. Falagas, E. A. Zarkadoulia, I. A. Bliziotis, and G. Samonis. Science in Greece: from the age of Hippocrates to the age of the genome. *FASEB J.*, 20(12):1946–1950, 2006.

[15] M. Fatemi, A. Manduca, and J. F. Greenleaf. Imaging elastic properties of biological tissues by low-frequency harmonic vibration. *Proceedings of the IEEE*, 91(10):1503–1519, Oct. 2003. 25

[16] I. Foster. *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering*. Addison-Wesley, Don Mills, ON, 1995. 98

[17] D. Fu, S. F. Levinson, S. M. Gracewski, and K. J. Parker. Non-invasive quantitative reconstruction of tissue elasticity using an iterative forward approach. *Phys Med Biol*, 45:1495–1509, Jun 2000. 8

[18] K. J. Glaser, J. P. Felmlee, and R. L. Ehman. Rapid MR elastography using selective excitations. *Magn Reson Med*, 55:1381–1389, Jun 2006.

[19] G. H. Golub. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 3rd edition, 1996. 57

[20] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, New Jersey, 3rd edition, 2008. 64

[21] M. A. Green, L. E. Bilston, and R. Sinkus. In vivo brain viscoelastic properties measured by magnetic resonance elastography. *NMR Biomed*, 21:755–764, Aug 2008.

[22] M. E. Haacke, R. W. Brown, M. R. Thompson, R. Venkatesan, M. E. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Liss, June 1999. 33, 44

[23] J. Hadamard. Sur les problèmes aux derivées partielles et leur significations physiques. *Princeton University Bulletin*, 13:1–20, 1902. 44

[24] L. Huwart, F. Peeters, R. Sinkus, L. Annet, N. Salameh, L. C. ter Beek, Y. Horsmans, and B. E. Van Beers. Liver fibrosis: non-invasive assessment with MR elastography. *NMR Biomed*, 19:173–179, Apr 2006.

[25] K. Ito, M. Kroller, and K. Kunisch. A numerical study of an augmented lagrangian method for the estimation of parameters in elliptic systems. *SIAM Journal on Scientific and Statistical Computing*, 12(4):884–910, 1991.

[26] F. Kallel and M. Bertrand. Tissue elasticity reconstruction using linear perturbation method. *IEEE Trans Med Imaging*, 15:299–313, 1996. 8

[27] T. Kariya. *Generalized Least Squares*. Wiley, New Jersey, 2004. 48

[28] D. Klatt, P. Asbach, J. Rump, S. Papazoglou, R. Somasundaram, J. Modrow, J. Braun, and I. Sack. In vivo determination of hepatic stiffness using steady-state free precession magnetic resonance elastography. *Invest Radiol*, 41:841–848, Dec 2006. 10

[29] H. Knutsson, C. F. Westin, and G. Granlund. Local multiscale frequency and bandwidth estimation. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 1, pages 36–40, Nov 1994. 10, 67

[30] S. A. Kruse, G. H. Rose, K. J. Glaser, A. Manduca, J. P. Felmlee, C. R. Jack, and R. L. Ehman. Magnetic resonance elastography of the brain. *Neuroimage*, 39:231–237, Jan 2008.

[31] D. S. Lake, A. Manduca, Y. Mariappan, and R. L. Ehman. Directional filtering for multiple-driver magnetic resonance elastography data. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 14, page 330, 2006. 135

[32] R. Luce and S. Perez. Parameter identification for an elliptic partial differential equation with distributed noisy data. *Inverse Problems*, 15(1):291–307, 1999.

[33] A. Manduca, R. Muthupillai, P. J. Rossman, J. F. Greenleaf, and R. L. Ehman. Local wavelength estimation for magnetic resonance elastography. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 527–530, Sep 1996. 10

[34] A. Manduca, R. Muthupillai, P. J. Rossman, J. F. Greenleaf, and R. L. Ehman. *Visualization of tissue elasticity by magnetic resonance elastography*, volume 1131 of *Lecture Notes in Computer Science*, pages 63–68. Springer Berlin, 1996. 10

[35] A. Manduca, V. Dutt, D. T. Borup, R. Muthupillai, R. L. Ehman, , and J. F. Greenleaf. *Reconstruction of Elasticity and Attenuation Maps in Shear Wave Imaging: An Inverse Approach*, volume 1496 of *Lecture Notes in Computer Science*, pages 606–613. Springer Berlin, 1998.

[36] A. Manduca, V. Dutt, D. T. Borup, R. Muthupillai, J. F. Greenleaf, and R. L. Ehman. Inverse approach to the calculation of elasticity maps for magnetic resonance elastography. In K. M. Hanson, editor, *Medical Imaging 1998: Image Processing*, volume 3338, pages 426–436. SPIE, 1998. 8

[37] A. Manduca, T. E. Oliphant, M. A. Dresner, J. L. Mahowald, S. A. Kruse, E. Amromin, J. P. Felmlee, J. F. Greenleaf, and R. L. Ehman. Magnetic resonance elastography: non-invasive mapping of tissue elasticity. *Med Image Anal*, 5:237–254, Dec 2001. 1, 6, 25

[38] A. Manduca, T. E. Oliphant, D. S. Lake, M. A. Dresner, and R. L. Ehman. Characterization and evaluation of inversion algorithms for mr elastography. In M. Sonka and J. M. Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684, pages 1180–1185. SPIE, 2002.

[39] A. Manduca, D. S. Lake, S. A. Kruse, and R. L. Ehman. Spatio-temporal directional filtering for improved inversion of MR elastography images. *Med Image Anal*, 7:465–473, Dec 2003. 67, 74, 119, 129, 130

[40] Y. Mariappan, P. J. Rossman, and R. L. Ehman. Magnetic resonance elastography with multiple drivers. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 9, page 617, 2005. 135

[41] Y. K. Mariappan, P. J. Rossman, K. J. Glaser, A. Manduca, and R. L. Ehman. Magnetic resonance elastography with a phased-array acoustic driver system. *Magn Reson Med*, 61:678–685, Mar 2009. 5, 6, 135

[42] R. Masuzaki, R. Tateishi, H. Yoshida, T. Sato, T. Ohki, T. Goto, H. Yoshida, S. Sato, Y. Sugioka, H. Ikeda, S. Shiina, T. Kawabe, and M. Omata. Assessing liver tumor stiffness by transient elastography. *Hepatol Int*, 1:394–397, Sep 2007. 1

[43] P. J. McCracken, A. Manduca, J. Felmlee, and R. L. Ehman. Mechanical transient-based magnetic resonance elastography. *Magn Reson Med*, 53:628–639, Mar 2005. 5, 10

[44] M. D. J. McGarry, H.-U. Berger, and E. E. W. V. Houten. Damping models in elastography. In A. Manduca and X. P. Hu, editors, *Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*, volume 6511, page 65111W. SPIE, 2007. 20

[45] R. Muthupillai and R. L. Ehman. Magnetic resonance elastography. *Nat. Med.*, 2:601–603, May 1996.

[46] R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman. Magnetic resonance elastography by direct visualization of propagating acoustic strain waves. *Science*, 269:1854–1857, Sept 1995. 5, 6, 9, 37

[47] R. Muthupillai, P. J. Rossman, D. J. Lomas, J. F. Greenleaf, S. J. Riederer, and R. L. Ehman. Magnetic resonance imaging of transverse acoustic strain waves. *Magn Reson Med*, 36:266–274, Aug 1996.

[48] J. Nocedal. *Numerical Optimization*. Springer, New York, 2nd edition, 2006. 49

[49] T. E. Oliphant. *Direct Methods for Dynamic Elastography Reconstruction: Optimal Inversion of the Interior Helmholtz Problem*. PhD in Biomedical imaging, Mayo Graduate School, Rochester, MN, 2001. 6, 11, 25, 76, 84, 119, 129

[50] T. E. Oliphant, J. L. Mahowald, R. L. Ehman, and J. F. Greenleaf. Complex-valued quantitative stiffness estimation using dynamic displacement measurements and local inversion of conservation of momentum. In *Ultrasonics Symposium, 1999. Proceedings. 1999 IEEE*, volume 2, pages 1641–1644, 1999. 11, 25, 60, 73

[51] T. E. Oliphant, R. L. Ehman, and J. F. Greenleaf. Magnetic resonance elastography revisited: reinterpreting phase-difference data as the output of a linear filter. In *Proceedings of the 8th Annual Meeting of ISMRM*, page 539, 2000. 39

[52] T. E. Oliphant, A. Manduca, A. Dresner, R. Ehman, and J. Greenleaf. Adaptive estimation of piece-wise constant shear modulus for magnetic resonance elastography. In *Proc. Intl. Soc. Mag. Reson. Med 9*, page 1642, 2001. 11

[53] T. E. Oliphant, A. Manduca, R. L. Ehman, and J. F. Greenleaf. Complex-valued stiffness reconstruction for magnetic resonance elastography by algebraic inversion of the differential equation. *Magn Reson Med*, 45:299–310, Feb 2001. 25, 74

[54] E. Park and A. M. Maniatty. Shear modulus reconstruction in dynamic elastography: time harmonic case. *Phys Med Biol*, 51:3697–3721, Aug 2006. 11, 26, 27, 62

[55] D. B. Plewes, J. Bishop, A. Samani, and J. Sciarretta. Visualization and quantification of breast cancer biomechanical properties with magnetic resonance elastography. *Phys Med Biol*, 45:1591–1610, Jun 2000. 5

[56] J. G. Proakis. *Digital Signal Processing*. Pearson Prentice Hall, New Jersey, 4th edition, 2007. 64

[57] A. J. Romano, J. J. Shirron, and J. A. Bucaro. On the noninvasive determination of material parameters from a knowledge of elastic displacements theory and numerical simulation. *IEEE Trans Ultrason Ferroelectr Freq Control*, 45:751–759, 1998. 10, 78

[58] I. Sack, B. Beierbach, U. Hamhaber, D. Klatt, and J. Braun. Non-invasive measurement of brain viscoelasticity using magnetic resonance elastography. *NMR Biomed*, 21:265–271, Mar 2008.

[59] R. Sinkus, J. Lorenzen, D. Schrader, M. Lorenzen, M. Dargatz, and D. Holz. High-resolution tensor MR elastography for breast tumour detection. *Phys Med Biol*, 45:1649–1664, Jun 2000. 12, 25

[60] R. Sinkus, M. Tanter, S. Catheline, J. Lorenzen, C. Kuhl, E. Sondermann, and M. Fink. Imaging anisotropic and viscous properties of breast tissue by magnetic resonance-elastography. *Magn Reson Med*, 53:372–387, Feb 2005. 12, 25

[61] R. Sinkus, M. Tanter, T. Xydeas, S. Catheline, J. Bercoff, and M. Fink. Viscoelastic shear properties of in vivo breast lesions measured by MR elastography. *Magn Reson Imaging*, 23:159–165, Feb 2005. 26

[62] A. R. Skovoroda, S. Y. Emelianov, and M. O'Donnell. Tissue elasticity reconstruction based on ultrasonic displacement and strain images. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 42(4):747–765, Jul 1995.

[63] A. R. Skovoroda, L. A. Lubinski, S. Y. Emelianov, and M. O'Donnell. Reconstructive elasticity imaging for large deformations. *IEEE Trans Ultrason Ferroelectr Freq Control*, 46:523–535, 1999. 5

[64] M. Stastna. Continuum mechanics notes. Lecture notes, January 2008. Department of Applied Mathematics, University of Waterloo. 21

[65] A. Tarantola. *Inverse Problem Theory and Methods for Parameter Estimation*. SIAM, Philadelphia, 2005. 57

[66] G. Tenti. Theory of viscoelasticity. Lecture notes, January 1994. Department of Applied Mathematics, University of Waterloo. 21

[67] L. N. Trefethen. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997. 49, 57

[68] E. E. Van Houten, K. D. Paulsen, M. I. Miga, F. E. Kennedy, and J. B. Weaver. An overlapping subzone technique for MR-based elastic property reconstruction. *Magn Reson Med*, 42:779–786, Oct 1999. 8, 98

[69] E. E. Van Houten, J. B. Weaver, M. I. Miga, F. E. Kennedy, and K. D. Paulsen. Elasticity reconstruction from experimental MR displacement data: initial experience with an overlapping subzone finite element inversion process. *Med Phys*, 27:101–107, Jan 2000. 9

[70] E. E. Van Houten, M. I. Miga, J. B. Weaver, F. E. Kennedy, and K. D. Paulsen. Three-dimensional subzone-based reconstruction algorithm for MR elastography. *Magn Reson Med*, 45:827–837, May 2001.

[71] E. E. Van Houten, M. M. Doyley, F. E. Kennedy, J. B. Weaver, and K. D. Paulsen. Initial in vivo experience with steady-state subzone-based MR elastography of the human breast. *J Magn Reson Imaging*, 17:72–85, Jan 2003. 9

[72] E. E. Van Houten, M. M. Doyley, F. E. Kennedy, K. D. Paulsen, and J. B. Weaver. A three-parameter mechanical property reconstruction method for MR-based elastic property imaging. *IEEE Trans Med Imaging*, 24:311–324, Mar 2005. 9

[73] M. Yin, J. A. Talwalkar, K. J. Glaser, A. Manduca, R. C. Grimm, P. J. Rossman, J. L. Fidler, and R. L. Ehman. Assessment of hepatic fibrosis with magnetic resonance elastography. *Clin. Gastroenterol. Hepatol.*, 5:1207–1213, Oct 2007.

[74] A. A. Young and L. Axel. Three-dimensional motion and deformation of the heart wall: estimation with spatial modulation of magnetization–a model-based approach. *Radiology*, 185:241–247, Oct 1992.

[75] Y. Zheng, Q. C. Chan, and E. S. Yang. Magnetic resonance elastography with twin drivers for high homogeneity and sensitivity. *Conf Proc IEEE Eng Med Biol Soc*, 1:1916–1919, 2006. 6