2D and 3D Elasticity Imaging Using Freehand Ultrasound

Joel Edward Lindop

Pembroke College

March 2008

A dissertation submitted to the University of Cambridge for the degree of Doctor of Philosophy.
Summary

Medical imaging is vital to modern clinical practice, enabling clinicians to examine tissues inside the human body non-invasively. Its value depends on accuracy, resolution, and the imaged property (e.g., density). Various new scanning techniques are aimed at producing elasticity images related to mechanical properties (e.g., stiffness) to which conventional forms of ultrasound, X-ray and magnetic resonance imaging are insensitive.

Elastography, palpography or strain imaging has been under development for almost two decades. Elasticity images are produced by estimating and analysing quasistatic deformations that occur between the acquisition of multiple ultrasound images. Likely applications include improved diagnosis of breast cancer (which often presents as a stiff lump), but the technique can be unreliable and difficult to perform. Practical imaging is based on freehand scanning, i.e., the ultrasound probe is moved manually over the surface of the tissue. This requires that elasticity images are calculated fast to provide a live display, and the images need to present meaningful elasticity data despite the poorly controlled properties of the deformations.

This thesis presents technical developments towards clinically practical elasticity imaging. First, deformation estimation is examined to devise algorithms that are both computationally efficient and accurate. Second, the entire image formation process is considered, providing strain data accompanied by indications of accuracy, which are then appropriately scaled and displayed in elasticity images representing the value and reliability of elasticity data.

Displacements are estimated by matching windows of radio-frequency data between pre- and post-deformation ultrasound frames. Robust tracking ensures that displacement estimates can be found by searching over small ranges without introducing large errors, location estimation corrects a well-known amplitude modulation artefact, and a “weighted phase separation” framework illuminates the scope for optimising the speed and accuracy of deformation estimators.

Strain estimates derived from each estimated deformation provide a form of elasticity image. A method is devised for predicting the accuracy of each strain estimate, which is first applied for dynamic resolution selection: parameters are automatically modulated to produce images with fixed precision at variable resolution. This indicates the scope for using accuracy indicators, which are applied to greater practical advantage in an interface concept: Nonuniform normalisation of strain data leads to “pseudo-strain” images. Values from multiple images are blended adaptively to produce a final display that is reliable, while indicating the level of uncertainty where data are less accurate.

This has made it possible to produce good 2D and 3D elasticity images by freehand scanning. Indeed, a clinical trial has recently been set up to evaluate the utility of this system in various clinical scenarios.

Keywords: medical imaging, 2D ultrasound, 3D ultrasound, freehand, elastography, elasticity imaging, strain imaging, deformation imaging, RF signal processing
Acknowledgements

I am grateful to Graham Treece for the guidance and imagination with which he supervised my research. I like to discuss things, and Graham is very patient, so we have chatted about elasticity imaging for literally hundreds of hours. Additionally, Graham maintained the Stradwin software used for acquiring all of the ultrasound data in this thesis. Thanks go also to Andrew Gee and Richard Prager for their advice at intervals throughout this research.

I was keen to try and make this project clinically relevant, so Lol Berman kindly hosted me in the ultrasound unit at Addenbrooke’s Hospital on several occasions. More recently, Sue Freeman has begun clinical work with this imaging system. She performed the scanning for several of the figures.

Patrick Gosling and Anna Langley did a great job of making sure my computer worked at critical moments. I also enjoyed Anna’s lute and mandolin playing, while Patrick regaled me on many an afternoon with his unique analyses covering topics from medical imaging, child-rearing and the Wall Street Crash to the Ashes 2005.

Ben Dickins, Matt Sangster and James Kemp proof-read some of the following chapters. I am grateful to several people for occasional discussions, some about ultrasound, others scientific or philosophical, and many miscellaneous, which may have contributed to the thinking behind this thesis: Pierre-Yves Baccou, Vee Barbary, Katie Birkwood, Tom Bishop, Francesco Bonaccorso, Lujie Chen, Ben Dickins, Tom Drummond, Peter Hassenpflug, James Housden, Oliver Lindop, James Ng, Vidura Seneviratne, Michal Sobótka, Deepak Subramanian, Richard West and Terry Wilson.

Finally, many thanks are due to my parents for their support throughout this process, and to Anastazja for support and encouragement over the past two years. She is going to make pierogi when I submit this thesis.
Declaration

This dissertation is the result of my own original work and does not include anything done in collaboration with others, except where specifically indicated in the text. It has neither been submitted in whole nor in part for a degree at any other university. It contains 88 figures and approximately 64,000 words including figure captions, footnotes, tables, equations, appendices and the bibliography.

The following publications have been derived from this work:

**Patent applications**


**Journal articles**


**Conference presentations**


# Contents

Glossary vii

I Introduction 1
  1.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
  1.2 Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
  1.3 Elasticity imaging concepts . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
    1.3.1 Quasistatic . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
    1.3.2 Dynamic (continuous) . . . . . . . . . . . . . . . . . . . . . . . . . 16
    1.3.3 Dynamic (transient) . . . . . . . . . . . . . . . . . . . . . . . . . . 21
    1.3.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
  1.4 Quasistatic elasticity imaging in 2D and 3D . . . . . . . . . . . . . . . . . 26
  1.5 Original contributions and thesis outline . . . . . . . . . . . . . . . . . . . 35

II Deformation estimation 37
  2 Robust tracking 38
    2.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
    2.2 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
      2.2.1 Tests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
      2.2.2 Tracking strategies . . . . . . . . . . . . . . . . . . . . . . . . . . 41
    2.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
    2.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
    2.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
  3 Location estimation 57
    3.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
    3.2 Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
      3.2.1 AM noise . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
      3.2.2 AM correction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
      3.2.3 AMC for phase zero methods . . . . . . . . . . . . . . . . . . . . . 63
## CONTENTS

3.2.4 AMC for correlation coefficient methods .................................................. 67
3.2.5 AM suppression alternatives ...................................................................... 68
3.3 Methods .............................................................................................................. 70
  3.3.1 Algorithms ................................................................................................ 70
  3.3.2 Simulation .................................................................................................. 73
  3.3.3 Phantom ..................................................................................................... 73
  3.3.4 In vitro and in vivo scanning ..................................................................... 74
3.4 Results ............................................................................................................... 74
  3.4.1 Window length ............................................................................................ 74
  3.4.2 Compression factor .................................................................................... 75
  3.4.3 Strain dependence ...................................................................................... 76
  3.4.4 Phantom study ........................................................................................... 76
  3.4.5 In vitro and in vivo results ......................................................................... 81
3.5 Discussion ......................................................................................................... 84
3.6 Conclusion ........................................................................................................ 86

4 Phase-based estimators ....................................................................................... 87
  4.1 Background ..................................................................................................... 87
  4.2 Theory .............................................................................................................. 87
    4.2.1 Weighted phase separation ..................................................................... 88
    4.2.2 Weighting selection ............................................................................... 90
  4.3 Methods .......................................................................................................... 94
    4.3.1 Algorithms ............................................................................................... 94
    4.3.2 Tests ......................................................................................................... 95
4.4 Results .............................................................................................................. 96
4.5 Discussion ......................................................................................................... 101
4.6 Conclusion ....................................................................................................... 104

III Image formation ............................................................................................... 105

5 Dynamic resolution selection ........................................................................... 106
  5.1 Background .................................................................................................. 106
  5.2 Development ................................................................................................ 107
    5.2.1 Resolution .............................................................................................. 107
    5.2.2 Strain estimation error ............................................................................ 109
    5.2.3 Displacement estimation error ................................................................. 110
    5.2.4 Importance of the strain level ................................................................. 114
    5.2.5 Dynamic Resolution Selection procedure ........................................... 116
  5.3 Experiments .................................................................................................... 118
    5.3.1 Image formats ......................................................................................... 119
    5.3.2 Simulations .............................................................................................. 119
    5.3.3 Resolution .............................................................................................. 119
Glossary

\( \lambda \) the wavelength of ultrasound at the centre frequency

1D/2D/3D “one-/two-/three-dimensions”, “one-/two-/three-dimensional”, etc.

6DOF six degrees of freedom, i.e., arbitrary rigid-body motion in 3D

AMC amplitude modulation correction (see Chapter 3)

anechoic not echogenic; does not give rise to echos

anisotropic not isotropic; direction-dependent

ARFI acoustic radiation force impulse

artefact a misleading image feature

ASE adaptive strain estimation

B-scan a conventional ultrasound image; a 2D image showing the amplitude/envelope of ultrasound signals

centre frequency the centroid of the frequency spectrum (of an ultrasound signal)

decorrelate become less similar

decorrelation reduced similarity (between pre- and post-deformation ultrasound signals)

displacement change in position

downsampling discarding samples to reduce the sampling rate

DRS dynamic resolution selection (see Chapter 5)

echogenic gives rise to echos

echogenicity the strength with which ultrasound is reflected by a region of tissue containing scatterers

EPZS efficient phase zero search (see Chapter 3)

ex vivo (scanning) of excised tissue

FDA U.S. Food and Drug Administration; the body responsible for regulating medical equipment in the United States

FEM finite element modelling

freehand free manual control (of an ultrasound probe)
GLOSSARY

heterogeneous containing multiple regions with different properties
HIFU high intensity focused ultrasound
hydrophone device for recording sound in liquids, similar to a microphone
hypoechoic relatively weakly echogenic
inhomogeneities locations where physical properties change
insonification transmission of ultrasound (into a region of tissue)
in vitro (scanning) of laboratory samples
in vivo (scanning) of tissue in a living body
isotropic the same in all directions; direction-independent
MA moving average
MRI magnetic resonance imaging
off-line not real-time
orthotropic (anisotropic) having different properties defined along orthogonal directions
phantom an object fabricated for testing ultrasound scanning
PLL-SR piecewise-linear least squares regression
PWDE point-wise displacement estimate
real-time (computations) with real-world time constraints (such as needing to process ultrasound data as they are acquired)
ROI region of interest
RF radio-frequency
scatterers small inhomogeneities that reflect ultrasound
SNR signal-to-noise ratio
SNR strain estimation signal-to-noise ratio
speckle random-looking, fine-grained amplitude variation caused by interference between signals from closely-spaced sources
strain measures of deformation excluding translation and rotation based on spatial derivatives of displacement
subsample between samples; a fraction of the sample spacing
transducer any device converting signals between different physical bases (e.g., from ultrasonic pressure waves to electrical signals)
ultrasonic relating to ultrasound
ultrasound pressure waves at high frequencies ($\gg 20$ kHz)
WPS weighted phase separation (see Chapter 4)
Part I
Chapter 1

Introduction

1.1 Background

There is a long history of clinical practice in which the mechanical properties of tissue have been considered significant for medical diagnosis. In “The Book of Prognostics” (c. 400 BC) Hippocrates wrote regarding abdominal swellings that,

“Such, then, as are painful, hard, and large, indicate danger of speedy death; but such as are soft, free of pain, and yield when pressed with the finger, are more chronic than these.” [1] (emphasis added)

The tactile properties of tissue continue to represent important information for modern medical practitioners. Huge investment has been directed at research and infrastructure for breast screening programmes in developed countries [2, 3, 4]. Irrespective of the movement towards evidence-based medicine, with great emphasis being placed on measuring tangible changes in medical outcomes, manual palpation in the clinical breast examination is still widely regarded as an important procedure, contributing to a lowering of the breast cancer mortality rate [5]. On the other hand, breast cancer also exemplifies the limitations of subjective examinations when the tools may be inadequate. For many years it was assumed that training women in self-examination of their own breasts would achieve better medical outcomes, owing to earlier cancer detection, but the accumulated evidence indicates that such training programmes have only one significant result: The rate of biopsies on benign lesions goes up, which may in fact be damaging to health [6].

The limitations of manual examinations, looking for hard lumps, motivate the development of more advanced diagnostic tools. All forms of diagnosis have a subjective element, but hopefully new technology will help by providing clinicians with much richer information than is available through their fingertips. Physicists have long been interested in the mechanical properties of biological tissues [7, 8], but for obvious practical reasons this complex topic has attracted less attention than the study of engineering materials, so little exists by way of tabulated data that would be relevant to medical diagnosis [9]. However, such quantitative measurements as have been performed indicate that changes in stiffness may differentiate between healthy and diseased tissue in many cases [9, 10, 11, 12]. To
CHAPTER 1. INTRODUCTION

exploit such differences with a diagnostic imaging technique, this thesis focuses on the technical development of a practical system for producing images related to mechanical properties by means of freehand ultrasound scanning. This may be a useful approach to both 2D and 3D imaging of various tissues, building on the existing concept of quasistatic elasticity imaging [13, 14].

This work is motivated by general medical interest in mechanical property imaging, and specific advantages associated with quasistatic elasticity imaging, which can be appreciated in the context of the alternatives that are under development elsewhere. A broad introduction is provided, considering a range of related ultrasonic techniques, in which superficial differences mask common principles. Some aspects of the signal processing methods developed for “2D and 3D elasticity imaging using freehand ultrasound” may be applicable to a range of other imaging concepts.

The key to assessing mechanical properties is the measurement of motion, for which ultrasound is highly suitable. Clinicians have been able to observe movement in scans since the advent of real-time ultrasound in the 1980s. Working with image sequences, Dickinson and Hill demonstrated a correlation method for measuring small motions within tissue [15]. Subsequently, numerous researchers have proposed ultrasonic methods for assessing mechanical properties, with wide variation in the physical principles involved [13, 16, 17, 18, 19, 20, 21]. All methods involve the in vivo application of reversible deformations to human tissue, so imaging techniques based on measurement and analysis of such elastic deformations are called elasticity imaging.

Concurrently, other researchers proposed methods using magnetic resonance imaging (MRI) [22, 23]. One of the significant advantages of ultrasound is the relative ease of producing a real-time display, which enables clinicians to scan investigatively and concentrate on regions of interest. In general, conventional approaches to MRI and ultrasound have relative merits that are well known and likely to carry over to elasticity imaging. Ultrasound machines are extremely portable and far cheaper. MRI has the advantage of producing 3D scans of large volumes, while achieving fairly isotropic image quality. Ultrasound can match or surpass the resolution of MRI in the “axial” direction, but the quality of ultrasound images is highly anisotropic. Methods for generating focused ultrasound beams are called beamforming. Conventional beamforming leads to poorer resolution in the “lateral” and “elevational” directions (see Figure 1.1) [27]. Future advances may improve the non-axial resolution [28]. Anyway, there are various methods for performing 3D ultrasonic imaging with conventional beamforming [29, 30, 31], and 3D elasticity imaging is among the goals of this thesis. In the long run, successful techniques for elasticity imaging developed with MRI and ultrasound may be complementary in the context of general clinical practice.

Ultrasonic elasticity imaging is likely to provide valuable information for a wide range of clinical applications. Breast imaging to detect and diagnose cancer is a huge driver of research, and may be among the first applications to enter routine clinical practice.

In another recent development, compact ultrasound machines are available commercially as add-ons for laptop computers [24, 25, 26].
Figure 1.1: Naming convention for three orthogonal directions relative to an ultrasound probe. The “axial” direction is normal to the face of the probe (vertical in a 2D ultrasound image), and the “lateral” direction is parallel to the transducer array (horizontal in a 2D ultrasound image). These are both orthogonal to the “elevational” direction (out-of-plane with respect to a 2D ultrasound image).

[32, 33, 34, 35, 36, 37, 38]. Other studies have looked at diagnosing prostate cancer [39, 40], which causes similar changes in mechanical properties. Further applications are likely to emerge considering other soft tissues. A commercial system has already been released for grading liver fibrosis by means of ultrasonic elasticity measurements [41]. Elasticity imaging systems modified for dermatological scanning may improve skin cancer diagnosis [42, 43]; chronic dermatological conditions including scleroderma and Ehlers-Danlos syndrome are also characterised by changes in skin elasticity [44]. Several cardiovascular diseases are known to cause changes in mechanical properties. Elasticity imaging may improve the detection and staging of deep vein thrombosis [45], and the detection of vulnerable plaque deposits in patients with atherosclerosis [46, 47, 48]. Similar information would be useful in examinations of the myocardium [49, 50]. Further non-diagnostic applications include monitoring the formation of thermal lesions during ablation therapy (a form of non-invasive surgery) [51, 52, 53]. Techniques providing real-time images may also have intra-operative uses, such as improving the delineation of tumour boundaries during neurosurgery [54].

1.2 Theory

An overwhelming number of formally defined mechanical properties could be considered in the analysis of human tissue. Examples associated with elastic deformations are listed in Table 1.1. There is likely to be a subset of properties for which measurements or images would be clinically useful, and another subset of properties for which measurement or imaging is to some extent possible. The overlap is the type of property that may serve as the basis for successful elasticity imaging. An overview of relevant theory is presented,
CHAPTER 1. INTRODUCTION

Table 1.1: Examples of formally defined mechanical properties.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Names</th>
<th>Symbols</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear elasticity</td>
<td>Elastic modulus tensor</td>
<td>$C_{ijkl}$</td>
<td>A $3 \times 3 \times 3 \times 3$ matrix with up to 21 independent coefficients, describing 3D linear-elastic behaviour of <em>anisotropic</em> materials.</td>
</tr>
<tr>
<td></td>
<td>Lamé coefficients</td>
<td>$\mu$, $\lambda$</td>
<td>Often used in physics, these coefficients describe <em>isotropic</em> linear elasticity.</td>
</tr>
<tr>
<td></td>
<td>Shear modulus, bulk modulus</td>
<td>$G$ ($G = \mu$), $K$ (often $K \simeq \lambda$)</td>
<td>Referred to in engineering; an equivalent description of isotropic linear elasticity.</td>
</tr>
<tr>
<td></td>
<td>Young’s modulus, Poisson’s ratio</td>
<td>$E$, $\nu$</td>
<td>An equivalent description with appeal to intuitive understanding.</td>
</tr>
<tr>
<td>1D linear viscoelasticity</td>
<td>Creep compliance, relaxation modulus</td>
<td>$J(t)$, $Y(t)$</td>
<td>Functions of time indicating strain and stress responses to step changes in stress and strain respectively.</td>
</tr>
<tr>
<td></td>
<td>Complex compliance</td>
<td>$G_1(\omega) + jG_2(\omega)$</td>
<td>A complex representation indicating the magnitude and phase of strain divided by stress as a function of frequency in tissue subject to harmonic loading.</td>
</tr>
<tr>
<td>Linear isotropic poroelasticity</td>
<td>Young’s modulus, Poisson’s ratio, permeability</td>
<td>$E_s$, $\nu_s$, $k$</td>
<td>The Kwan-Lai-Mow biphasic model of porous media: these parameters describe an isotropic linear-elastic matrix saturated with fluid.</td>
</tr>
</tbody>
</table>

framing the role of qualitative approaches to elasticity imaging.

When analysing ultrasound images to estimate mechanical properties, the type of motion required is deformation, *i.e.*, compression, expansion, or shear. This is most simply described by a field of displacement data, recording tissue motion as a function of spatial position. A more useful description is in terms of strain, *i.e.*, quantities calculated by taking spatial derivatives of the displacement field to remove components associated with rigid body motion (bulk translation and rotation) which are unrelated to mechanical properties.

In 1D, strain is typically defined as the change in length divided by either the original
Figure 1.2: Components of 2D strain, $\varepsilon_{ij}$, when a square (solid line) is deformed (dashed line): (a) Longitudinal strain ($i = j$) indicates shortening or lengthening in a particular direction. (b) Shear strain ($i \neq j$) indicates warping as shown.

or the final length \[^7\]. Definitions differ significantly when it comes to large strains (e.g., greater than 10\%). Various definitions are suitable for 3D analysis, of which the most popular is also the simplest, and they converge when considering small deformations. Throughout this thesis, “strain” means elements from Cauchy’s strain tensor:

$$\varepsilon_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \quad i, j = 1, 2, 3, \quad (1.1)$$

where $u_i$ is displacement in direction $i$, and $x_j$ is pre-deformation position in direction $j$.

The different meanings of longitudinal and shear strain are indicated in Figure 1.2.

Strain arises due to changes in the stress field (force per unit area) within the tissue, which in turn follows changes in forces acting internally or on boundaries, called “mechanical excitation”. The tissue may be squashed by a compression plate \[^13\], vibrated from its surface \[^16\], struck on the surface by a small mass \[^20\], or palpated internally by radiation force \[^59\] (see Section 1.3.2). The resulting stress field comprises (1) an isotropic/hydrostatic component causing change in volume (change in pressure), and (2) anisotropic components causing change of shape (shear stresses and anisotropic components of longitudinal stresses). Elements from the 3D stress tensor are labelled in Figure 1.3.

In order to estimate mechanical properties, measurements or estimates of deformation must be combined with prior knowledge, estimates, or reasonable assumptions regarding mechanical excitation. In theory, deformation and excitation are linked via constitutive equations dependent on mechanical properties of the tissue. For example, the theory of linear elasticity based on Hooke’s law has been developed extensively by mathematicians, physicists and engineers. It has proven useful for analysing the properties of man-made structures such as buildings and machines \[^55\]. Much of the behaviour of common engi-

\[^1\]For small deformations, the distinction between pre- and post-deformation position is insignificant.
Figure 1.3: Elements from the 3D stress tensor labelled on an infinitesimally small cube. Force components acting on the surface of a finite object can be found by integration, e.g., \( dF_{11} = \sigma_{11} dx_2 dx_3 \). Strains with matching subscripts (not labelled above) indicate matching deformations, so the tissue has stretched over direction 1 if \( \varepsilon_{11} \) is positive, while a positive value of \( \varepsilon_{23} \) indicates warping in the 2–3 plane along the lines of Figure 1.2b.

Engineering materials, such as steel and timber, is actually poorly described by linear elasticity, but the theory is valuable because engineers are familiar with situations in which it applies relatively accurately [60].

The usual application for constitutive equations is to calculate deformation given knowledge of mechanical excitation and material properties: the “forward problem”. Working in reverse, it is sometimes possible to calculate the values that must apply regarding material properties given that a known mechanical excitation causes a known (measured) deformation: the “inverse problem”. This can often be solved, but inverse problems are more difficult than forward problems [61].

If a technique is required for producing quantitative images of a specific mechanical property by solving an inverse problem, one of the associated challenges is the need for accurate deformation measurements, which is considered at length throughout this thesis. Realistically, computational cost is also a significant issue, i.e., the time required to produce each image, and the amount of computing power that must be built into the ultrasound scanner. Both considerations apply similarly with regard to producing images that are only qualitatively related to mechanical properties.

Quantitative imaging presents three other fundamental challenges: (1) Complexity: This is perhaps the greatest obstacle. The mechanical behaviour of tissue is not accurately described by simple, convenient models [7]. (2) Computational stability: As the number of parameters in a model increases, the chance of a unique solution being available diminishes, and the quantity of data required for achieving acceptably low error increases dramatically [61]. (3) Unknown boundary conditions: Elasticity imaging is usually based on scanning a limited region of tissue, spanning a small volume (3D scans not encompassing the entire human body) or a slice (2D scans). This causes problems, because some approaches to quantitative analysis only produce correct solutions if the mechanical
properties are known all over the boundary \[62\].

For example, linear elasticity might seem the obvious constitutive equation for analysing biological tissue. Assuming fairly slow deformation, so that inertial terms are insignificant, the equilibrium equation is

\[
\sigma_{ij} = \sum_k \sum_l C_{ijkl} \varepsilon_{kl}, \quad i, j, k, l = 1, 2, 3, 
\]

(1.2)

where \(\sigma\) is the stress tensor, \(C\) is the elastic modulus tensor, and \(\varepsilon\) is the strain tensor. Symbols \(i, j, k\) and \(l\) denote the three spatial dimensions. Stresses and strains are longitudinal where \(i = j\) or \(k = l\), while shear stresses and shear strains have subscripts \(i \neq j\) or \(k \neq l\). The mechanical properties described by \(C\) may vary throughout the tissue. Equation 1.2 can be used in the forward problem to calculate the stress and strain fields throughout a region with known \(C\) subject to known boundary forces. The inverse problem can be tackled by calculating strain fields based on trial sets of mechanical properties and boundary forces. Likely parameter sets are identified by assessing the degree of correspondence between the predicted and measured strain fields. Usually the identification of probable parameter sets is also guided by prior assumptions regarding their relative plausibility \[61\]. When the most probable solution has been adopted, coefficients in \(C\) can be mapped to pixel values in quantitative elasticity images.

This example is intended to indicate the value of circumspection when deciding how to analyse mechanical properties for elasticity imaging. Linear-elastic theory is appealing because it enables the use of familiar linear algebra techniques \[63\], while accepting the possibility of real tissue being highly anisotropic. The mechanical properties of every piece of tissue are described in \(C\) by 21 independent coefficients \[55\]. Consequently, static analysis requires reasonably accurate assumptions regarding the 21-dimensional mechanical properties over the boundary, otherwise much of the model is redundant. Concerning computational stability, a single static deformation measured with perfect knowledge of the boundary conditions and perfect estimates of every strain component throughout the scan region is insufficient to yield a unique solution. Any chosen solution depends on prior assumptions \[61\]. It may be possible to identify a unique maximum likelihood solution, independent of prior knowledge, if several different deformations are analysed together. However, the inverse problem may remain ill-conditioned: the most likely solution may contain enormous errors even for data that are almost noise-free \[63\]. Prior knowledge remains important for enforcing a plausible solution.

Linear-elastic analysis may nevertheless be a good approach to elasticity imaging. Prior knowledge for practical analysis could include assumptions of dependent relationships between some of the coefficients in \(C\), so as to reduce the dimensionality of the problem. There is little point in adopting a highly complex linear-elastic model if tissue nonlinearity limits its validity, and time-dependent effects in the loading response introduce further errors. A diverse range of coefficients can be measured to describe tissue in detail when analysing \textit{ex vivo} samples \[7, 9\], but comprehensive characterisation of mechanical properties is unlikely to be a practical basis for elasticity imaging. The anal-
CHAPTER 1. INTRODUCTION

ysis needs to focus on salient properties of interest to clinicians, accepting that neglected properties may contribute to errors and artefacts.

Basic biophysics motivates a range of strategies. Soft tissues in the human body are complex composites. They consist mostly of fluid-filled sacks (cells), which are almost incompressible, i.e., the volume does not change, but cells present almost no resistance to shear [7]. Structural integrity is provided by an extra-cellular network of fibres, of which there are numerous types. Elastin exhibits highly linear-elastic behaviour, but is only a minor constituent except in the skin and vasculature. The most significant structural material is collagen, consisting of long, helical, covalently-bonded protein molecules, twisted like rope. This is extremely difficult to analyse (see Chapter 6 of [7]). The microstructure of the fibre network determines its macroscopic properties, which are usually anisotropic and nonlinear [9]. Friction associated with reorganisation of fibres contributes to viscous behaviour in most biological materials. At the time of writing, it seems sensible to make fairly basic approximations so as to realise practical imaging techniques, while the development of convenient models that more accurately characterise the mechanical properties of tissue is an evolving research topic [8].

A minority of elasticity imaging concepts emphasise the value of measuring viscous behaviour. Stress in viscoelastic tissue depends on the rate of deformation as well as its size. This is probably a significant factor in the behaviour of all soft tissues [7]. It has been suggested that measurements of viscoelastic time constants would relate to biochemical changes, providing information otherwise inaccessible by ultrasonic methods [64]. This is an exciting concept, but viscoelastic modelling is extremely complicated. Even 1D theory places no limit on the number of parameters that may be required for an accurate description of loading responses over time, depending on which type of linear model best matches the tissue behaviour [7, 56]. Researchers in ultrasound [65] and MRI [66] have suggested that the simplest viscoelastic model (the Kelvin-Voigt solid [56], which is rather unlike tissue [7]) can be applied where viscosity would otherwise be neglected, so as to improve the accuracy of quantitative images depicting other elastic moduli. This sort of development might be useful, although the increase in accuracy has not been measured [65, 66].

A more direct application of viscoelastic analysis is “poroelasticity imaging”. Fluid in tissue is rarely free to flow, except in vessels. However, it can do so in cases of oedema, where fluid accumulates and causes swelling. Fluid flow associated with elevated porosity can be assessed qualitatively using images based on the change in non-axial strain over time in tissue subject to a static uniaxial load [67, 68]. It may be possible to extract meaningful estimates of parameters to describe poroelastic behaviour quantitatively, assuming that the tissue consists of a solid, isotropic, linear-elastic matrix, saturated with incompressible, near-inviscid fluid, characterised by a permeability coefficient that relates pressure gradients to flow rates [58] (see Table 1.1). Whether or not progress is made towards quantitative poroelasticity imaging [69], preliminary studies suggest that the qualitative images may be useful in themselves for assessing oedematous tissue [68, 69].

By contrast, the most common basis for elasticity imaging is a simplification that ig-
nores viscosity. The pragmatic theoretical motivation follows the quotation in Section 1.1. Salient differences between hard and soft tissues have long been detected by touch through manual palpation. It is anticipated that elasticity imaging techniques targeting the same physical property may help by offering superior accuracy and/or better spatial resolution. Against the complexity described above, the mechanical properties of tissue are reduced to a single parameter, assuming the behaviour to be isotropic, linear-elastic, inviscid and incompressible.

Can these assumptions be justified? Firstly, while tissue is rarely isotropic, the scope for extracting more information by adding to the model complexity may be limited, since deformations are often influenced mainly by the mechanical properties in a single direction. Secondly, although tissue behaviour is highly nonlinear under large deformations, any deformation can be linearised if it is sufficiently small.† There are three common ways of expressing the assumed constitutive equation under equilibrium conditions [55]:

\[
\sigma_{ij} = \begin{cases} 
\lambda \sum_k \varepsilon_{kk} + 2 \mu \varepsilon_{ij} & \text{if } i = j, \\
2\mu \varepsilon_{ij} & \text{otherwise}, 
\end{cases}
\]

(1.3a)

\[
\sigma_{ij} = \begin{cases} 
K \sum_k \varepsilon_{kk} + 2G (\varepsilon_{ij} - \frac{1}{3} \sum_k \varepsilon_{kk}) & \text{if } i = j, \\
2G \varepsilon_{ij} & \text{otherwise}, 
\end{cases}
\]

(1.3b)

\[
\sigma_{ij} = \begin{cases} 
\frac{E}{1+\nu} (\varepsilon_{ij} + \frac{\nu}{1-2\nu} \sum_k \varepsilon_{kk}) & \text{if } i = j, \\
\frac{E}{1+\nu} \varepsilon_{ij} & \text{otherwise}. 
\end{cases}
\]

(1.3c)

\(\lambda, \mu\), \((K, G)\) and \((E, \nu)\) are the parameter pairs associated with each formulation (see Table 1.1). Equation 1.3a is the most elegant mathematical expression, but Equation 1.3b is easier to explain. Change in volume multiplied by \(K\) gives the isotropic component of stress, and anisotropic components of strain multiplied by 2\(G\) give anisotropic stress components.‡ The parameters in Equation 1.3c are also widely referred to, despite the cumbersome 3D formulation. (For uniaxial stress, \(E\) is the ratio between axial stress and strain, while \(\nu\) is the ratio between non-axial and axial strain.)

In any event, these formulations are equivalent. Each parameterisation separates components of elastic behaviour in a useful way if tissue is assumed to be incompressible. This would be an unreasonable assumption for porous tissue, as discussed above, but it applies fairly accurately in most healthy tissue, provided that fluid channels such as blood vessels and lymph ducts are not a large fraction of the tissue volume. This fixes one of the elasticity parameters \((\lambda \simeq \infty, K \simeq \infty, \text{ or } \nu = 0.5)\), so the remaining parameter \((\mu, G, \text{ or } E)\) fully characterises any variation. It has been suggested that this accounts almost entirely for the useful information accessed by manual palpation [10, 70].

It may be possible to apply a small load to a region of tissue, measure the deformation, and analyse the strain pattern to infer variation in the shear modulus, \(G\), or Young’s modulus, \(E\).§ Several factors lend ambiguity to this, not the least of which is viscosity.

---

†“Sufficiently small” is an unknown quantity that requires clarification through practical experience.
‡The factor of 2 owes to the different definitions of “engineering stress” and \(\varepsilon_{ij}\) from Equation 1.1.
§Incompressibility implies \(E = 3G\).
Equation 1.3 concerns static loads, whereas in reality all loads are dynamic to some extent. Viscous damping may have a significant impact if the loading rate is too high, or viscous creep may be significant if static loads are applied for a long time (minutes or hours) [7]. Somewhere between those extremes is a “quasistatic” loading rate at which dynamic effects are genuinely unimportant, hence “quasistatic elasticity imaging”.

Alternatively, tissue elasticity can be investigated by measuring and analysing dynamic behaviour. If viscous effects are insignificant, differences between quasistatic and dynamic behaviour arise due to inertia. Keeping with assumptions of isotropic linear elasticity, Newton’s second law of motion gives [55]

$$\rho \frac{\partial^2 u}{\partial t^2} = \left( K + \frac{4}{3} G \right) \nabla(\nabla \cdot u) - G \nabla \times \nabla \times u,$$

where $\rho$ is density. This is really the superposition of two wave equations, one for isotropic strain, the other for shear strain. The pressure wavespeed, $c_p$, and the shear wavespeed, $c_s$, are

$$c_p = \sqrt{\frac{K + \frac{4}{3} G}{\rho}}, \quad \text{and} \quad c_s = \sqrt{\frac{G}{\rho}}.$$

Conventional ultrasound images (B-scans) show the amplitude of pressure waves in the ultrasonic frequency range reflected back from inhomogeneities in soft tissue. This is a successful imaging technique because $K$ and $\rho$ are often substantially constant, so the speed of sound, $c_p$, exhibits little variation. As a result, the depth of a reflector can be considered directly proportional to the arrival time of its echo after a pressure pulse has been transmitted from the ultrasound probe into the tissue.

All forms of ultrasound image are distorted if the speed of sound varies, although there is some scope for correcting minor distortion [71]. When $c_p$ is constant, it is possible to acquire raw ultrasound data very quickly to record the progress of shear waves propagating through the tissue. The assumption that soft tissue is incompressible implies $c_p \gg c_s$, because $K$ is very large. For example, in fat the pressure wavespeed is roughly 1000 times greater than the shear wavespeed [72]. Theoretically, up to 500 B-scans can be acquired in the time taken for a shear wave to propagate over the length of an image. It is therefore possible to devise ultrasonic methods for estimating shear wavespeed, which is related to the shear modulus by Equation 1.5. In practice, nonlinearity corrupts the shear waveform, while viscosity modifies the wavespeed and causes attenuation. Nevertheless, information provided by successful estimation of shear wavespeeds will in all likelihood have great clinical value.
1.3 Elasticity imaging concepts

The following review is not exhaustive, but it should serve as sufficient context for judging the significance of the approach pursued in this thesis. The main ultrasonic concepts are summarised, and differences are highlighted, considering three categories based on the time-dependency of mechanical excitation: quasistatic, dynamic (continuous), and dynamic (transient) [73].

1.3.1 Quasistatic

The basic scanning procedure consists of (1) recording a “pre-deformation” ultrasound image of unloaded tissue, (2) applying a load, and (3) recording a “post-deformation” ultrasound image [13]. The deformation between pre- and post-deformation ultrasound frames is estimated by a suitable signal processing technique, and analysed to produce an elasticity image.

That abstract covers numerous related techniques. Loading can be applied at a quasistatic rate by various means. For example, research into intravascular elasticity imaging exploits physiological excitation: artery walls deform because of the change in blood pressure over the cardiac cycle [74]. The loading has a frequency of approximately 1 Hz, which seems suitable for quasistatic analysis. In other tissues, quasistatic deformation arises due to breathing.

Many investigations have been performed using motion of the ultrasound probe as the source of quasistatic mechanical excitation. Tissue compresses when the probe is pushed firmly against the surface, and relaxes when the probe is held lightly (see Figure 1.4). Suppose that motion of the ultrasound probe causes a uniform change in the axial component of longitudinal stress. Strain at each point in the image is then inversely proportional to Young's modulus if the behaviour is isotropic linear-elastic. The stress field is usually nonuniform, so strain data are ambiguous, but strain imaging is the simplest way of displaying quasistatic deformation data to provide a visual indication of variation in mechanical properties. Techniques exploiting probe movement for mechanical excitation divide into two subcategories. Automated scanning requires extra hardware [13, 32], whereas freehand scanning is the usual technique employed in conventional sonography [14, 33, 39]. The choice of scanning technique has significant consequences.

Automated scanning entails mounting an ultrasound probe on a mechanical actuator to deform tissue by means of a carefully defined movement at the surface. For example, tissue can be compressed by translating the probe precisely 1 mm in the axial direction [13, 32, 75]. The main advantage is that a certain probe movement is repeatable, especially if it is known to lead to good elasticity images in a particular application. Furthermore, the reliable acquisition of suitable pre- and post-deformation ultrasound data means that computations for producing an elasticity image can be performed off-line. This might buy time for elaborate computational methods to maximise the quality of the elasticity images.
Figure 1.4: The principle behind quasistatic elasticity imaging: (a) Inhomogeneous tissue is modelled as a set of springs, where different spring constants represent variation in Young’s modulus. Bold springs are three times stiffer than the rest. An ultrasound image records the pre-deformation state of the tissue. (b) Tissue deforms when the ultrasound probe is pressed firmly against the surface, and a post-deformation ultrasound image is recorded. The deformation here is exaggerated for illustrative purposes (25% compressive strain). It is likely that smaller deformations (strain on the order of 1%) are more appropriate when examining human tissue. (c) Tissue displacement relative to the ultrasound probe is estimated by analysing the ultrasound images, and strain is calculated by differentiation. If the stress throughout the tissue is uniform (as in this example), strain is inversely proportional to the spring constant, so the strain in bold springs is three times lower than elsewhere. (d) Variation in the spring constant is revealed by setting pixel intensities according to the strain values.

Equally, the need for correct prior knowledge of a suitable probe movement is potentially a disadvantage. The best movement for quasistatic elasticity imaging is likely to vary from one patient to the next. Automated scanning is cumbersome compared to the freehand approach that clinicians are accustomed to. Usually the sonographer holds an ultrasound probe in his or her hand and moves it manually over the patient’s skin to produce images of accessible tissue from whichever angle proves to be the most informative \cite{33, 39}. There is inevitably some variation in the pressure applied through the probe between consecutive ultrasound frames in a freehand scan \cite{76}, which is a source of continuous mechanical excitation.

The freehand scanning approach is hugely appealing, because it can potentially be implemented on conventional ultrasound machines with minimal modification to the hardware and scanning procedure. Elasticity imaging could be made available as a display option to switch on whenever required during routine sonography. Such a development
might resemble the previous adoption of Doppler techniques for measuring blood flow [77, 78, 79], which has rapidly become an indispensable adjunct to conventional ultrasound imaging [27].

Compared to automated scanning, the deformation in a freehand scan is uncontrolled and unpredictable. Clinicians may be unable to perform prescribed probe movements accurately on a very fine scale. Therefore, freehand deformations may often be too small or too large for inferring mechanical properties, or the type of deformation may simply be unsuitable. Figure 1.4 shows a compression in which the probe is translated axially. Freehand scanning is bound to throw up more complicated movements, such as rotation about the elevational axis. Some freehand strain images must inevitably be less informative than automated strain images. On the other hand, a sequence of freehand strain images can test a range of deformations, so the best freehand images may be more informative than the best automated strain images [33]. Furthermore, the variation within a sequence of freehand deformations could increase the reliability of inferences relating to mechanical properties.

The value of a strain image sequence must depend on the sonographer’s scanning technique, which determines the deformation sequence. A live (real-time) strain display is likely to enhance the benefits of freehand scanning significantly by providing continuous feedback, so the sonographer can adjust his or her scanning technique towards whatever seems to work best [33]. Live strain displays have been demonstrated in the past with frame rates similar to conventional ultrasound imaging (some tens of Hz) [14, 80]. However, joint requirements for accuracy and speed place pressure on hardware and algorithm design.

Whether by automated or freehand scanning, quasistatic elasticity imaging may depict in fine detail the geometry of regions distinguished by their mechanical properties. The fundamental limit on resolution is the same as that of the ultrasound images from which the deformation estimates are derived. However, the usefulness of the images will depend on how accurately deformation can be estimated using ultrasound, and how closely inferences based on the deformation estimates can be made to correspond to quantitative mechanical properties of the tissue such as Young’s modulus.

The first point is critical, because raw displacement estimates are spatially differentiated when calculating strain or inferring mechanical properties, which tends to amplify noise. The signal strength can be increased by applying larger deformations, but very large deformations are unlikely to be desirable. Aside from the basic issue of patient comfort, the response is more nonlinear, and hence more difficult to interpret, as discussed in Section 1.2. Anyway, large deformations may not improve the signal-to-noise ratio, because an increase in the “deformation signal” causes greater decorrelation between successive ultrasound frames, which in turn increases the level of estimation noise [81]. This topic is considered in detail throughout this thesis. The feasibility of accurately estimating tissue deformation using ultrasound data is a basic premise behind this work, although there may be no simple means of ensuring that all deformation estimates are sufficiently accurate.
On the second point, while nonuniformity in the stress field means that strain images along the lines of Figure 1.4d are not equivalent to Young's modulus images, can strain data be converted into Young's modulus images? The inverse problem is already greatly simplified by the assumption of isotropic, incompressible, linear-elastic behaviour introduced in Section 1.2. Accepting the approximate nature of this analysis, it is possible to evaluate Young's moduli throughout the scan region if the stresses or Young's moduli are known at all points over the scan boundary [62]. Several researchers have attempted to produce quantitative images on this basis [82, 83, 84, 85, 86], but unknown boundary conditions are a significant obstacle. It is (probably) impossible to measure all of the boundary stresses, so assumptions are needed regarding boundary values of Young's modulus. It has been suggested that images of relative Young's modulus can be produced by assuming uniformity over the boundary [83, 85], but dramatic errors can occur if the assumption of uniformity is incorrect [62]. The selection of appropriate prior assumptions is emphasised in model-based approaches, which may be more successful for specific well-constrained tasks [87].

Challenges associated with the inverse problem should not impede simpler approaches to quasistatic elasticity imaging. Strain images represent useful information despite variation in stress. Boundaries in accurate strain images usually correspond to boundaries with respect to mechanical properties, although the size of differences in strain does not match the size of differences in Young's modulus [88]. Strain images cannot justify quantitative statements such as, “Young's modulus in the lesion is 2.5 times higher than in the background,” but it may often be reasonable to make qualitative observations such as, “The round lesion (diameter of 3 mm) is much stiffer than the background.” This point of interpretation is clarified in Section 1.4.

To date, strain imaging has been the approach in almost every clinical trial of quasistatic elasticity imaging. In one instance, data were acquired by automated scanning [32], but freehand scanning has been tested more frequently [33, 34, 35, 36, 37, 38, 39, 40] with a focus on breast and prostate examinations. Successful strain images are sometimes better than conventional ultrasound for identifying tumours [40]. Malignancy and different types of cancer may be distinguishable on the basis of either strain contrast [32] or (perhaps more likely) differences in geometrical appearance between conventional ultrasound and strain images [32, 33, 35, 36]. Reports conclude unanimously that strain imaging could potentially be a useful clinical tool, but none has yet demonstrated a conclusive case for adoption into routine clinical practice. The issue of stress variation and boundary conditions seems not to be a major drawback, because the qualitative information is still useful. The question of how to produce accurate images is more of a problem, because strain images continue to be unreliable [32, 33, 36]. The presence of many bad images makes interpretation difficult and laborious for sonographers, who are typically required to inspect sequences of stored strain images by eye to select those that are informative [36]. This thesis contributes towards improving the reliability of strain images.
1.3.2 Dynamic (continuous)

Can better elasticity images be produced by techniques based on continuous dynamic mechanical excitation? Usually a vibrator acts on the tissue surface, transmitting shear waves in the same frequency range as audible sound. Techniques with this form of excitation are known as **sonoelasticity imaging**. Alternatively, an oscillatory force in the audible frequency range can be applied to a small region within the tissue using high-intensity focused ultrasound, to which the mechanical response is measured in **vibro-acoustography**.

**Sonoelasticity imaging**

Lerner and Parker *et al.* [16, 89] pioneered the adaptation of Doppler methods to generate vibration-amplitude images showing patterns of steady-state vibration in soft tissue. Vibration at a low frequency (10–1000 Hz) is applied at the surface with an amplitude of roughly 0.1 mm. The typical imaging setup is shown in Figure 1.5a. Vibrators need to be installed at suitable locations to transmit shear waves throughout the scan region.

Example images are shown in Figure 1.6. Stiff inclusions in soft tissue show up as dark regions of low vibration amplitude. There are two main reasons for this. Firstly, a continuous travelling wave of a given intensity has lower amplitude inside an inclusion of high shear modulus. Secondly, reflections at inclusion boundaries accentuate the difference in vibration amplitude by reducing the wave energy that enters the inclusion, while the level of energy in the background is slightly increased [90].

![Figure 1.5: Setup for sonoelasticity imaging: (a) Vibration-amplitude imaging: One or more vibrators transmit shear waves into the tissue, and the amplitude is measured using Doppler ultrasound. (b) Crawling-wave imaging: Vibrators with near-identical frequencies transmit shear waves from opposite sides of the ROI, driving a modal pattern from which the shear wavespeed can be inferred. (c) Holographic-wave imaging: The ultrasound probe oscillates at a similar frequency to the vibrator. The relative motion in the tissue therefore appears as a slow-moving interference pattern, from which the shear wavespeed can be estimated. A thick layer of gel is required to couple ultrasound into the tissue without introducing extra shear waves owing to the probe motion.](image-url)
CHAPTER 1. INTRODUCTION

Figure 1.6: Vibration-amplitude images, reproduced by permission of Kenneth Hoyt, University of Rochester, NY, USA: (a) B-mode image showing a tissue-mimicking phantom with a stiff spherical inclusion, 7 mm in diameter, of matched echogenicity. There is almost no contrast between the inclusion and the background. (b) Vibration-amplitude image produced by mechanical excitation at 150 Hz. (c) Vibration-amplitude image produced by mechanical excitation at 250 Hz. Modal patterns dominate the 150 Hz image, so the 250 Hz image gives a better indication of the inclusion geometry.

The most obvious advantage of sonoelasticity imaging compared to quasistatic elasticity imaging is the robustness and simplicity of the signal processing. Vibration amplitude is estimated by a pulsed-Doppler method: A narrow-bandwidth pulse of ultrasound is transmitted into the tissue. Vibration amplitudes at various depths along the beam are estimated by windowing the echoes and calculating the spectral variance [91]. Fast computation and sufficient accuracy is provided for this task by the well-known autocorrelation method [92], first described by Kasai et al. [79]. Relatively low accuracy is tolerable because, unlike other elasticity imaging concepts, there is no need to take a derivative, and the level of noise is insignificant compared to vibration-amplitude artefacts.

The link between vibration amplitude and stiffness is indirect. Just as stress nonuniformity is a limitation of quasistatic strain images, so vibration-amplitude images are affected by nonuniform wave intensity owing to diffraction, refraction, reflections and viscous losses (wave intensity decays away from the vibration source). Viscous losses are frequency-dependent, so lower frequencies penetrate deeper into the tissue, but the images then have lower resolution [93]. There is no prospect of converting vibration-amplitude data into quantitative estimates of mechanical properties [90], but images such as Figure 1.6c indicate that qualitative interpretation of vibration-amplitude images could sometimes be useful for lesion detection.

Modal patterns are a significant additional artefact in vibration-amplitude images. They arise because of interference between steady-state shear waves travelling in opposite directions [92], so regions with elevated or reduced vibration amplitude appear in patterns unrelated to local changes in mechanical properties. The patterns are determined by the overall geometry of the scan target, its global mechanical properties, and the position and frequency of the vibrator. For example, the vibration-amplitude image in Figure 1.6b is
CHAPTER 1. INTRODUCTION

dominated by modal patterns. Comparison with Figure 1.6c shows that the geometry and strength of modal patterns depend on frequency. The artefacts can be made less severe by vibrating at multiple frequencies simultaneously; some modal patterns cancel out in the combined vibration-amplitude image, although they cannot be eliminated entirely [73, 92, 94].

As a consequence, vibration-amplitude images are probably more difficult to interpret than quasistatic strain images. Additionally, the practical scanning setup is more complicated. Limited progress has been made towards clinical demonstration: in vitro scanning of prostate tumours after excision indicated the possibility of producing useful images of relevant biological tissue [95, 96], but the prospect of vibration-amplitude imaging entering routine clinical practice seems remote.

Other types of sonoelasticity imaging might be more successful. A more complicated and noisier extension to the signal processing for vibration-amplitude imaging recovers phase as well as amplitude. Maps of amplitude and phase have been used by Yamakoshi et al. to construct images of wavefront propagation through a ROI [17]. Whereas vibration-amplitude images were intended for identifying small stiff inclusions [16, 89, 92], wavefront images are used for estimating shear wavespeed by a manual image-analysis procedure [17]. Successful estimates would be useful for assessing diffuse conditions in large organs such as the liver.

The manual approach to wavefront analysis serves as a noise rejection strategy, because users learn to ignore data that seem unreliable [97]. In practice, the final output is a single number (shear wavespeed) rather than an image. The technique has been evaluated in a clinical trial involving more than 200 liver patients [97]. Significant differences were recorded between the average wavespeed estimates from patients with different conditions, but there was surprisingly wide variation in the speeds measured within each group. Complicated effects associated with continuous vibration, including modal patterns, introduce bias into wavespeed estimates [20]. The manual approach to image analysis is extremely laborious, so the wavespeed estimates need to be very accurate and useful to justify the effort. Despite further technical development reported in [98], the technique seems finally to have been abandoned.

The most promising form of sonoelasticity imaging actually exploits modal patterns in vibration-amplitude images. It has been shown that special interference patterns can be generated to support shear wavespeed estimation. Two vibration sources of equal frequency produce a standing wave with nodal spacing proportional to the shear wavespeed, or a pair of vibrators at slightly different frequencies generates a moving interference pattern (the “crawling wave”) with speed proportional to the shear wavespeed multiplied by the difference frequency [99]. Alternatively, the ultrasound probe is vibrated at a frequency near that of the shear waves, which converts steady-state shear excitation from a single vibrator into a slow-moving interference pattern (the “holographic wave”) [100]. Figure 1.5b–c shows the setup for these techniques. Analysis of the interference patterns potentially leads to quantitative shear wavespeed images.

The major advantage of the crawling- and holographic-wave techniques compared
to the method of Yamakoshi et al. [17, 97] is that larger quantities of more accurate wavefront data are produced, which is a more promising base for developing automatic analysis methods [99, 100, 101, 102]. The absolute values of shear wavespeed estimates for homogeneous phantoms are accurate and precise [99, 102]. However, the holographic-wave technique is subject to bias as described by [20] if wavespeed is measured in inhomogeneous media. The crawling-wave technique only provides a correct estimate of shear wavespeed when the interfering wavefronts are planar and parallel [101], which is difficult to achieve in general, and may be impossible in inhomogeneous tissue. That limitation may yet be overcome by improving the analysis [101]. On the other hand, even if the absolute values are biased, the qualitative discrimination of a stiff inclusion has been demonstrated several times with in vitro scanning [100, 101, 102] (see Figure 1.7). In view of the complicated experimental setup, it will be interesting to see if either crawling- or holographic-wave imaging can be applied successfully to in vivo scanning.

Vibro-acoustography

The absorption of energy near the focus of a high-intensity focused ultrasound (HIFU) beam supplies heat and induces a small force along the beam direction, known as radiation force [103]. In some situations the force applies almost entirely within a small region near the focus [59]. There are several concepts for elasticity imaging exploiting this effect, mostly in the dynamic (transient) category (see Section 1.3.3). However, vibro-acoustography uses continuous excitation by radiation force, and is usually considered a form of elasticity imaging [70].
Figure 1.8: Setup for vibro-acoustography. A pair of confocal HIFU transducers at slightly different frequencies produce an oscillatory force at the difference frequency at their focus, from which sound is emitted and detected by a hydrophone. By scanning different points throughout the tissue, an image can be constructed of the sonic amplitude, which is related in a complicated way to mechanical properties.

The tissue is excited by means of a pair of confocal HIFU beams, transmitting at slightly different frequencies (see Figure 1.8). The foci overlap in a tightly confined region, where radiation force oscillates at the beat frequency (typically $\sim 10$ kHz). The resulting motion around the focus produces an acoustic wave detected by a hydrophone [19, 104]. An image is constructed by scanning over the ROI and plotting the measured power of the sound from each focal position. Pixels in a vibro-acoustogram represent the combined effects of variation in tissue properties above the focus (shadowing and refraction), the level of absorption at the focus, the size of the resultant oscillatory motion (which depends on mechanical properties near the focus), and the transfer function to the signal recorded by the hydrophone (which depends on bulk mechanical properties).

Therefore, vibro-acoustograms are entirely determined by mechanical properties. The images are free of speckle, which is advantageous compared to conventional ultrasound images, and they have high resolution, probably superior to that of any other elasticity imaging concept [19, 104, 105, 106]. However, the information provided by vibro-acoustography is very different to quasistatic elasticity imaging. It would be extremely complicated to relate pixel values in a vibro-acoustogram to a particular mechanical property, such as Young’s modulus. Vibro-acoustography may be a good method for detecting microcalcifications in breast tissue; the principle has been demonstrated by ex vivo scanning of histological samples [105, 106]. However, this thesis is concerned with producing images that exhibit contrast between regions of tissue that are hard or soft, which vibro-acoustograms do not [107]. Equally, conventional ultrasound images depend on mechanical properties, but a clinician cannot distinguish between hard and soft tissue by looking at a conventional ultrasound display.
1.3.3 Dynamic (transient)

Transient concepts involve applying impulsive mechanical excitation, to which tissue has a “transient” (short duration) response. Wherever the transient excitation is applied, stresses propagate into the surrounding tissue at a finite rate dependent on the wavespeed. In the first moments after excitation, the response depends solely on mechanical properties at the point of application.

The realisation of transient methods is technically challenging, because data need to be captured very rapidly. Ultrafast ultrasound scanning has been a significant technical development, employing novel beamforming to acquire 2D ultrasound frames at rates up to 10,000 Hz [108]. This is at least 100 times faster than current commercial scanners, so many 2D snapshots can be acquired to record the propagation of an individual shear wave travelling through soft tissue.

Transient shear imaging

Catheline et al. proposed transient elasticity imaging with impulsive excitation as a solution to biases in shear wavespeed estimates based on continuous excitation [20]. It was found that a suitable shear wave impulse with a centre frequency of \( \sim 100 \) Hz propagating along the axis of a single-element ultrasound transducer could be generated by vibrating the transducer itself [109]. A practical system was constructed by mounting the transducer on a vibrator in a handheld device [109] (see Figure 1.9). Accurate estimation of shear wavespeed in homogeneous tissue was demonstrated by \textit{in vivo} scanning.

Further efforts were made to produce 2D elasticity images by tracking the motion of a plane wave in 2D using ultrafast ultrasound scanning [110]. As with the single-element device, a shear wave can be generated by mounting the ultrasound probe on a handheld vibrator [111]. However, compared to the basic utility of shear wavespeed measurements characterising large tissue regions, the added value associated with 2D images from this

![Figure 1.9: Setup for using FibroScan® to measure shear wavespeed in the liver. The probe consists of a single ultrasound transducer mounted on a vibrator, which is held on the tissue surface transmitting ultrasound and shear waves though a gap between ribs. The shear wavespeed in liver is estimated over a 1D ROI, away from the tissue surface to avoid confusion with the wavespeed in subcutaneous tissue.](image-url)
CHAPTER 1. INTRODUCTION

Figure 1.10: The FibroScan® display [41], reproduced by permission of Laurent Sandrin, EchoSens, Paris, France (http://www.echosens.com). These images from clinical scans of liver in vivo show strain data over depth (vertical axis) against time (horizontal axis). The gradient of the dark stripe is the axial component of the shear wave velocity, from which stiffness may be inferred. (a) Healthy patient. (b) Mild fibrosis. (c) Liver cirrhosis.

device is apparently limited, because it is difficult to realise surface-generated shear waves that propagate sufficiently parallel to the imaging plane throughout the ROI to produce accurate images with good resolution [21, 111].

On the other hand, the simple scanning technique producing accurate, repeatable measurements of shear wavespeed has proven highly attractive to clinicians, even without images. The single-transducer device exists as a commercial product called FibroScan® [41]. This produces a single shear wavespeed estimate to characterise bulk stiffness in a large region of tissue, which must be affected by a diffuse condition (so that the change in stiffness is uniform) near to the surface (so that shear waves propagate sufficiently far through the ROI). Figure 1.10 shows examples reproduced from the first clinical study [41]. The liver has proven to be an ideal target, so FibroScan® has progressed rapidly towards widespread adoption as a tool for assessing liver fibrosis, with thousands of patients already included in clinical trials [112, 113, 114, 115, 116, 117].

Acoustic radiation force impulse (ARFI) imaging

Images related to mechanical properties can be constructed by analysing the response to impulsive excitation by radiation force [18]. A “pushing pulse” lasting a few ms causes sudden deformation near its focus, followed by a period of relaxation as tissue recovers its undeformed shape. This response depends mostly on the mechanical properties near the focus. ARFI is described as a natural successor to manual palpation: “radiation force of ultrasound is used instead of fingers” [18] for “remote palpation” [118]. Recent growth of interest in ARFI is partly due to the fact that the latest diagnostic ultrasound probes
can be driven at high enough power to generate pushing pulses and capture image data almost simultaneously [118].

The maximum displacement [118, 119] or the strain [120] is estimated at the ARFI focus, scanning over a 2D ROI to form an image. Suppose that the focal stress is uniform in magnitude (independent of material properties), applying across a small region of fixed size: The strain and the peak transient displacement at the focus are then inversely proportional to Young’s modulus, before stress propagates into the surrounding tissue [120]. In principle, it may be interesting to measure other characteristics of the ARFI response, such as time constants related to viscosity [119], but maximum displacement and strain have been considered more often.

Maximum displacement and strain do not yield quantitative Young’s modulus images. The size of the ARFI response depends on the level of absorption at the focus, and the level of attenuation in tissue above the focus [121, 122]. ARFI images are therefore qualitative in the same sense as quasistatic strain images, but with a different set of artefacts. Displacement imaging has been demonstrated in numerous ex vivo and in vivo situations [52, 123], exhibiting typical artefacts such as shadows, while nonetheless demonstrating that displacement images can highlight stiff regions in real tissue.

If ARFI is to be used widely for clinical purposes, an important development strand for safety is quantification of the heating that accompanies radiation force.† The instantaneous intensity required for ARFI is orders of magnitude higher than the FDA limit for average insonification intensity in diagnostic ultrasound scanning [121]. It also exceeds the FDA limit on peak intensity [21]. The pulses have short duration, so the heating and risk of cavitation can probably be restricted to safe levels by careful control of the scanning procedure, but this needs to be demonstrated convincingly. Palmeri et al. [125] modelled heating for various types of ARFI, indicating focusing methods and frame rates that should avoid temperature increases greater than a couple of °C. The predictions are subject to large uncertainty, however, owing to complex effects including nonlinearity, which can significantly change the level of absorption at the focus [103].

Assuming that a practical technique is shown to be safe, it would be useful to extend ARFI towards quantitative elasticity imaging. It is unlikely that ARFI images showing maximum displacement or strain at the focus can be converted reliably into estimates of mechanical properties. Once again, shear wavespeed estimation seems the most promising approach to acquiring quantitative data. Shear wavefronts could be tracked as they propagate away from the focus after each pushing pulse [59]. Nightingale et al. [122] conducted a proof-of-principle investigation looking at ex vivo and in vivo tissue. Two example images showed that meaningful measurements are possible, although the level of accuracy was poor. A low starting amplitude at the focus and high attenuation make it difficult to track the wavefronts [122].

†A more common application of HIFU is ablation therapy, i.e., burning tissue! [124]
Supersonic shear imaging

Supersonic shear imaging (SSI) was first described by Bercoff et al. [21, 53]. This method for shear wavespeed measurement based on ARFI involves sufficient novelty to merit separate consideration. ARFI is applied at a series of foci in quick succession, scanning axially up or down a column next to the ROI. The focus moves faster than the shear wavespeed, so a plane wave propagates away from the excitation column along the edge of a Mach cone. Ultrasound data acquired by ultrafast ultrasound scanning record the shear wave propagation. Sufficient data for calculating an entire shear wavespeed image are acquired in the time taken for a single plane wave to cross the ROI. Methods for extracting shear wavespeed estimates from these data have been discussed by McLaughlin and Renzi [126, 127].

SSI reportedly has several advantages [21]: The superposition of shear waves from different foci boosts the amplitude for a given intensity of insonification. Consequently, the demonstration system operated within FDA limits for average and peak intensity. Another advantage is the short time in which a plane wave crosses the ROI (around 30 ms for a 40×40 mm image) so frame rates above 10 Hz are possible. Finally, the orientation of the plane waves can be controlled by changing the speed and direction of scanning the excitation column. This can be exploited in a form of multi-orientation compounding to suppress noise.

Judging by published results, SSI can produce meaningful 2D sets of shear wavespeed estimates, but its resolution and repeatability are unclear [21, 53]. The dearth of publications may reflect the strategy of a new commercial venture: Supersonic Imagine, Aix-en-Provence, France. An SSI ultrasound machine is being developed for freehand scanning, with a single probe both providing the special pattern of radiation force and performing ultrafast ultrasound scanning. One of the challenges with SSI is the vast quantity of ultrasound data that must be processed to generate each shear wavespeed image. Real-time imaging might be achievable using dedicated computing hardware for highly parallel processing, but the likely significance of SSI is difficult to assess. The substantial investment attracted by Supersonic Imagine suggests that unpublished results may be encouraging. Either way, it seems reasonable to anticipate that, even if an SSI product is successful, the hardware requirements of the technique may be too demanding for it to become a common feature on general-purpose ultrasound machines.

†http://www.supersonicimagine.fr
1.3.4 Summary

Apart from vibro-acoustography, the concepts described above all target more or less the same material property, distinguishing between soft and hard tissues. Multiple techniques might be clinically useful if they offer advantages in different situations. Some will prove redundant, but there almost certainly is a role for quasistatic elasticity imaging by freehand scanning.

A distinction can be drawn between concepts that are either relatively qualitative or relatively quantitative. The former (quasistatic elasticity imaging, vibration-amplitude imaging, and ARFI) measure the size of the response to mechanical excitation, which depends partly on local mechanical properties, as well as on confounding factors, such as the local strength of the mechanical excitation (local stress, local shear wave intensity, or local radiation force). The latter concepts (crawling- and holographic-wave forms of sonoelasticity imaging, surface transient methods such as FibroScan®), and SSI) estimate shear wavespeed, which depends on local mechanical properties at each point in the tissue.

Quantitative data measured on an absolute scale are obviously valuable, because the numbers may in themselves be useful for diagnosis. However, the most accurate quantitative data may not correspond to the most informative elasticity images, because it would be useful to visualise the geometry of regions with different mechanical properties. The appearance of structure within an image will sometimes be more important than numerical accuracy in the correspondence between image contrast and Young’s modulus variation.

Even shear wavespeed has limitations as a quantitative measure. Where tissue is anisotropic, the meaning of shear wavespeed is potentially ambiguous. Equation 1.5 neglects viscous effects, which modify the shear wavespeed and make it frequency-dependent. Each piece of tissue still has a particular shear wavespeed for a particular polarisation and frequency. However, in addition to sources of inaccuracy mentioned in Sections 1.3.2 and 1.3.3, shear wavespeed estimates produced by 1D or 2D methods are inevitably subject to bias, since they neglect off-axis or out-of-plane components of shear wave velocity. The absolute values are only useful when the bias is small.

Of the approaches to measuring shear wavespeed, the FibroScan® technique and SSI are much simpler to perform than crawling- or holographic-wave forms of sonoelasticity imaging. They may also be more accurate, because the wavespeed estimates are more direct. Of the two, FibroScan® is significant, because it is already in widespread clinical use for measuring liver stiffness. However, SSI seems a more promising approach to producing accurate shear wavespeed images of inhomogeneous tissue.

If SSI has the potential to produce quantitative data in images that reveal geometric features, why bother with the qualitative concepts? Firstly, there may be situations in which one of the qualitative concepts produces the best combination of high resolution and low noise. SSI analyses the propagation of low-frequency shear waves with wavelengths of several millimetres. It would be surprising if the resolution of SSI did not partly depend

---

1The speed of shear waves that are polarised within a particular plane depends on the mechanical properties in the normal direction [128].
on the shear wavelength, so finer detail might be resolved by other concepts where fewer factors limit the resolving power. Secondly, SSI is likely to require relatively expensive hardware. Other techniques might be suitable for implementing on a wider range of hardware platforms at lower cost. Thirdly, considering that the performance of SSI is still unproven (at least in the public domain), progress with the other concepts will be all the more important if SSI performs less well than expected in practice.

Of the qualitative concepts, vibration-amplitude imaging seems the least promising in terms of image quality, although it may be useful for examining tissues that are inaccessible by other forms of mechanical excitation. When available, quasistatic elasticity and ARFI images are easier to interpret. Of the two, ARFI is probably more restricted in terms of resolution, because it is affected by the size of the focus of the radiation force. Generally, quasistatic elasticity and ARFI imaging might be complementary, given the different artefacts in the images, although quasistatic elasticity imaging is significantly more convenient. Unlike ARFI, safety is not an issue in relation to quasistatic elasticity imaging, and the hardware requirements are relatively modest.

Quasistatic elasticity imaging can be implemented with minimal modification if any to standard ultrasound hardware. There is almost nothing to be lost by adding new signal processing algorithms for a technique that might become clinically important. The freehand approach to quasistatic elasticity imaging has already been successfully demonstrated in real clinical applications, providing images of relatively high resolution in the context of elasticity imaging. It seems probable that forms of quasistatic elasticity imaging will become available on large numbers of general-purpose ultrasound machines.

Quasistatic elasticity imaging is not without limitations. It is likely to be most successful when the ROI is fairly shallow, because the size of deformation due to quasistatic changes in the surface load decays with depth. Having accepted that the images are unlikely to provide quantitative data except in special cases, the main difficulty with quasistatic elasticity imaging is the demands placed on skill and judgment of the clinician to produce and identify good images. A simple, easy, robust system is needed if the potential clinical benefits are to be realised. This engineering problem is tackled by this thesis.

1.4 Quasistatic elasticity imaging in 2D and 3D

An approach is taken based on strain imaging with subtle modifications. Some variation in stress is adjusted for by variable scaling of measured strain values to produce “pseudo-strain” images. This is not an attempt to produce Young’s modulus estimates. The aim is to augment strain images by means of heuristic corrections, so that pseudo-strain data exhibit fewer artefacts and are easier to interpret.

This section examines the fundamental basis for strain and pseudo-strain imaging, clarifying the extent to which strain or pseudo-strain images may be informative. Results from a preliminary investigation conducted by the author are reviewed [14], providing the rationale for the developments in this thesis.

\[\text{Beta versions of strain imaging have been released on top-end ultrasound machines by Hitachi Medical Corporation (Tokyo, Japan) [34] and Siemens Medical Solutions (Malvern, PA, USA) [38].}\]
Fundamental basis

How far does stress variation distort the features in strain images produced by freehand scanning? Simulated strain fields for known mechanical properties indicate the sort of results that should be achievable by accurate strain or pseudo-strain imaging. Finite element modelling (FEM) was applied to simulate deformations of inhomogeneous tissue when a rigid block (the ultrasound probe) presses against the upper surface. Figures 1.11–1.14 show results in five cases calculated using commercial FEM software (ABAQUS 6.7 from SIMULIA, Providence, RI, USA).

Figure 1.11a shows the simulated geometry. Tissue is represented by a block of non-rigid material, with two inclusions (one hard, one soft) in background material of medium stiffness. The probe presses against the upper surface of the tissue, which rests on a rigid plate. The simulations are 2D, assuming “plane stress”, treating the out-of-plane dimension as 100 mm. The in-plane dimensions of the tissue are 80 × 60 mm, both inclusions are 10 mm in diameter, and the probe is 40 mm wide. Contact between the tissue and the rigid blocks is frictionless. Deformation occurs when the probe moves down 0.6 mm in the axial direction† against the upper surface of the tissue.

The simulated tissue is near-incompressible and linear-elastic. Case one consists of isotropic material with a Poisson’s ratio of 0.49. Compared to the background, Young’s modulus is three times higher in the hard inclusion and three times lower in the soft inclusion. Images of the resulting axial strain, axial stress and lateral strain are shown in Figure 1.11b–d. The axial strain image yields an intuitive representation of the mechanical properties in this simple case. The inclusions show up in the axial strain image with sharp boundaries and high contrast. There is very little lateral stress; the lateral strain image resembles the axial strain image owing to incompressibility.

The presence of undeformed regions on either side of the probe is inconsequential, because these would not fall within the ultrasound image. Strain in the background material is not entirely uniform, despite uniform Young’s modulus. The effects of stress concentrations are especially visible at the probe edges‡. The stress image explains the nonuniform background strain. Stress contrast also highlights the inclusions — for this reason the strain image fails to provide quantitative stiffness data, with a ratio of just 3.74 between the axial strains at the centres of the two inclusions, despite a Young’s modulus ratio of 9.00.

Cases two and three in Figure 1.12 consist of the same background material, but the mechanical properties of the inclusions are anisotropic in a sense known as “orthotropy”: linear elasticity is described by three Young’s moduli (one for each orthogonal direction), three shear moduli (one for each orthogonal plane), and three Poisson’s ratios. Near-incompressibility only implies $\nu \simeq 0.5$ in isotropic material. [129] provides formulae for calculating the Poisson’s ratios of near-incompressible orthotropic material as functions of

---

† The horizontal, vertical and out-of-plane directions are referred to as lateral, axial and elevational as usual (cf., Figure 1.1).

‡ The perfectly sharp edges of the simulated probe exaggerate the size of the stress concentrations compared to those that would occur in reality.
Figure 1.11: FEM case one: simulated compression of an isotropic, linear-elastic medium containing one hard inclusion and one soft inclusion. (a) Illustration of geometry. (b) Axial strain (range black–white represents 0–2% compressive strain). (c) Axial stress (0–2% of the Young’s modulus of the background material). (d) Lateral strain (0–1% tensile strain).

Figure 1.12: Simulated compression of orthotropic, linear-elastic media containing one hard inclusion and one soft inclusion. Image greyscale mappings are the same as in Figure 1.11. (a) FEM case two: stiffness variation only in the axial direction. (a1) Axial strain. (a2) Lateral strain. (b) FEM case three: stiffness variation only in the lateral and elevational directions. (b1) Axial strain. (b2) Lateral strain.
the Young’s moduli. The inclusions in Figure 1.12a have the same lateral and elevational Young’s moduli as the background, but the axial Young’s modulus is three times higher in the hard inclusion and three times lower in the soft inclusion. Conversely, the inclusions in Figure 1.12b have the same axial Young’s modulus as the background, but the lateral and elevational Young’s moduli differ.

The inclusions in Figure 1.12a are more similar to the background in terms of their mechanical properties than the inclusions in Figure 1.11. However, the axial and lateral strain images in Figure 1.12a1–2 are remarkably similar to Figure 1.11b,d. On the other hand, the inclusions in Figure 1.12b are almost invisible in the strain images, although their Young’s moduli differ in two directions. The lesson from these cases is that strain images are sensitive to the component of variation in mechanical properties in the axial direction, and almost independent of the properties in the other directions. This is important, because many biological tissues are anisotropic. Additionally, all three cases indicate that lateral strain images may be superfluous for near-incompressible tissue, because they reveal almost no additional information beyond that provided by axial strain images.

Further cases could be presented with more complex properties such as nonlinearity, but this has little importance from the perspective of qualitative imaging. Materials may vary in stiffness along their loading curves, and strain images still differentiate between regions that are relatively stiff and relatively soft, providing an indication of their geometries. A greater concern is the extent to which extreme variation in stress within the scan region may sometimes produce misleading images. This is related to the issue of boundary conditions in quantitative analysis, because a region surrounded by material with uniform properties is likely to have a fairly uniform stress field at its boundary.

Case four in Figure 1.13 represents isotropic material in a 10 mm layer. Hard and soft inclusions are sandwiched directly between the compression plates. This scenario is not unlikely: a scan could be performed examining small nodules in a thin layer of soft tissue over bone. The probe moves down just 0.1 mm, so the average strain is the same as in the earlier cases. The inclusions are visible in the strain image because of contrast at their edges, but there is no contrast between the axial strains at their centres. This strain pattern could just as well be associated with stiffness variation in the material surrounding the inclusions as with the inclusions themselves. The critical information is the stress image: the ratio between the stresses at the centres of the inclusions is 7.11, which is almost as high as the Young’s modulus ratio.

Figure 1.13: FEM case four: simulated compression of a thin layer of an isotropic, linear-elastic medium containing one hard inclusion and one soft inclusion. Image greyscale mappings are the same as in Figure 1.11. (a) Axial strain. (b) Axial stress.
Figure 1.14: FEM case five: simulated uneven compression of an isotropic, linear-elastic medium containing one hard inclusion and one soft inclusion. The probe rotates about its right-hand edge, resulting in uneven compression. Image greyscale mappings are the same as in Figure[1.11] (a) Axial strain. (b) Axial stress.

The inclusions in Figure[1.13] would be detectable by manual palpation, because clinicians sense the pressure applied through their fingers. A non-ultrasonic form of elasticity imaging is being developed using a handheld palpation device with pressure sensors on its surface. Surface pressure measurements can be converted into a mechanical image [130]. Surface pressure and strain data are obviously complementary. Where the strain image is relatively poor, contact pressure may be more informative, and vice versa, as in Figure[1.11c]. There are many situations in which strain images are likely to be useful. An awareness should build with experience of the types of scan targets for which strain imaging is appropriate.

Extreme nonuniformity in the stress field can also arise due to uneven probe movement during a freehand scan. Case five in Figure[1.14] consists of the same material as Figure[1.11]. Now the probe pivots about its right-hand edge, so Figure[1.14b] shows low stress on the right and the original stress on the left, with gradual variation in between. This changes the brightness in different parts of the strain image in Figure[1.14a], but the boundaries between both inclusions and the background are still visible. Having accepted the qualitative nature of strain imaging, a diffusely varying strain pattern has little value for assessing tissue properties, because it could just as well be related to diffuse variation in stress.[†] In that case, little or no significant information has been lost going from Figure[1.11b] to Figure[1.14a].

Fundamentally, the FEM simulations indicate that accurate strain or pseudo-strain images from freehand scanning could provide useful information relating to mechanical properties. The feasibility of the technique rests on whether sufficiently accurate strain estimates can be produced by analysing the pre- and post-deformation ultrasound data. This is the primary technical challenge in quasistatic elasticity imaging. Deformation estimation for this purpose is somewhat more complicated than for other elasticity imaging concepts that involve dynamic mechanical excitation, because motion must be estimated on two levels:

[*Indeed, these simulations are 2D, with a uniform elevational dimension. In reality, stress decreases with depth, because it spreads out elevationally away from the imaging plane.*]
• Bulk component: Despite not considering large static deformations (all strains are \(\ll 10\%\)) some of the displacements may be large since they accumulate across the image. Accurate estimation of these bulk components is a basic requirement so that sections of matching tissue can be compared between pre- and post-deformation ultrasound frames to facilitate accurate estimation of motion on a fine scale.

• Fine component: High accuracy is critical in the estimation of fine-scale motion to construct the displacement field. High-frequency noise in displacement fields is amplified by differentiation when calculating strain.

A more common task in medical imaging is image registration based on the approximate estimation of bulk movement [131, 132]; the Doppler effect is used to estimate blood velocity components without identifying absolute displacements [77, 78, 79]; or “speckle tracking methods” are applied to estimate absolute displacement as the basis for velocity estimation, where large errors (proportional to the size of the motion) are acceptable [133]. Some dynamic elasticity imaging concepts require accurate estimates of local deformation, potentially without a bulk component, so accurate displacement estimates can be calculated by focusing on a very short search range [134]. This is not the case in quasistatic elasticity imaging, which involves these two levels of displacement estimation. In the literature there tends to be a division between work focusing on achieving the greatest possible accuracy irrespective of computational cost [135, 136, 137], set against other work on less accurate algorithms for quick deformation estimation, suitable for live elasticity displays during scanning [80, 138, 139]. This dichotomy may not be inevitable. Deformation estimation is considered throughout Part II of this thesis.

Past experience

A preliminary investigation into 2D and 3D elasticity imaging using freehand ultrasound was based on the same hardware platform as most of this thesis, so details are now provided [14]. This is an ongoing project adding elasticity imaging to the freehand 3D ultrasound system developed by the Medical Imaging Group at the University of Cambridge [31], which has a real-time interface for the acquisition of RF ultrasound data [140]. The preliminary investigation demonstrated the feasibility of a freehand approach to 2D and 3D strain imaging.

Figure 1.15 shows the physical layout of the freehand 3D ultrasound system, comprising a Dynamic Imaging Diasus ultrasound machine (Dynamic Imaging Ltd., Livingston, UK) from which RF ultrasound data are sampled by a Gage\(^{†}\) CompuScope 14200 analogue-to-digital converter, processed on a PC running Stradwin\(^‡\) software, and displayed in real time. An AdapTrax\(^§\) target attached to the ultrasound probe is tracked by a Northern Digital\(^††\) Polaris optical position sensor, providing accurate 3D position mea-

\(^†\)http://www.gage-applied.com
\(^‡\)http://mi.eng.cam.ac.uk/~rwp/stradwin
\(^§\)http://www.traxtal.com
\(^††\)http://www.ndigital.com
measurements for every ultrasound frame. Work on elasticity imaging has been performed using a 5–10 MHz probe, which in practice has a centre frequency of \( \sim 6.0 \) MHz.

Freehand 3D ultrasound scanning entails recording conventional 2D ultrasound images while manually moving the ultrasound probe to sweep out a volume, and measuring the 6DOF position of each scan plane. The pixels from all of the 2D images are positioned correctly in a 3D coordinate system, from which images can be constructed by volume rendering or selecting an arbitrary plane slicing through the volume [31].

The idea of freehand 3D elasticity imaging arose due to the success of “probe-pressure correction”. Sometimes shapes in freehand 3D ultrasound images appear distorted because objects deform during the sweep. The main reason for this deformation is variation in the pressure applied through the probe. A proven method for reducing this distortion consists of [76]: (1) performing non-rigid registration to estimate the within-plane deformation between each pair of consecutive 2D images; (2) accumulating deformation estimates from one 2D frame to the next to calculate the net deformation relative to a reference frame; and (3) warping every 2D frame to reverse the estimated deformation. This reduces distortion in 3D image reconstructions by removing relative deformation. Probe-pressure correction requires deformation estimation similar to quasistatic elasticity imaging. The required accuracy is lower, but the success of probe-pressure correction suggested that freehand 3D elasticity imaging might be feasible [76], which was subsequently demonstrated by the preliminary investigation [14].

The scanning protocol shown in Figure 1.16 involves translating the probe slowly in the elevational direction (e.g., 3 cm in 10 s while acquiring 300 ultrasound frames) so consecutive frames are sufficiently correlated for performing accurate deformation estimation. Compared to conventional freehand 3D ultrasound scanning, the main difference in
Figure 1.16: Scanning protocol for freehand 3D elasticity imaging. (a) The probe is slowly swept manually over the tissue surface, with mechanical excitation due to slight pressure variation during the sweep, which may be inadvertent or deliberately applied. (b) A side view shows how freehand 3D elasticity imaging works. Deformation can be estimated between consecutive frames provided there is sufficient overlap.

performing a 3D sweep for elasticity imaging is the need to ensure that the probe pressure does vary slightly, whereas usually sonographers attempt to minimise pressure variation. This is a subtle difference, because the required level of pressure variation is quite small, and small variation is often unavoidable. Axial strain is calculated from the axial displacement field associated with each pair of consecutive ultrasound frames. Compared to the typical strain imaging techniques described in Section 1.3.1, one of the main additional challenges with 3D imaging is that artefacts arise in 3D reconstructions because the size of the deformation varies between different 2D images. They need to be rescaled to a consistent dynamic range before combining in 3D images, which are referred to as “pseudo-strain” to indicate the use of variable scale factors.

Various tests were performed, including scanning a phantom with a stiff inclusion (see Figure 1.17a) [14]. The images in Figure 1.17b–i come from a single 3D sweep consisting of 390 2D images. A convention was followed of mapping high pseudo-strain values to low pixel intensity (soft tissue appears dark) and low pseudo-strain to high pixel intensity (hard tissue appears light). Figure 1.17b–g shows a selection of 2D strain images. In some the stiff inclusion is clearly visible, while others are dominated by noise. Preliminary tests were also performed considering basic 2D strain imaging. In both cases, the quality of 2D strain images from freehand scanning is mixed. There are many reasons for variation in quality, such as different sizes of deformation, and different levels of elevational motion, which causes decorrelation [142]. An orthogonal reslice through the volume is shown in Figure 1.17h after the images have been combined in a 3D reconstruction. One of the 2D images is displayed together with a couple of reslices in Figure 1.17i giving an outline of the 3D pseudo-strain volume. The position of the inclusion can be identified in 3D, because many of the 2D images are reasonably good, but the appearance is streaky, because there are also a lot of poor strain data.
Clearly, good 2D strain or pseudo-strain images can be produced by a freehand scanning approach. They can be viewed in a real-time display or incorporated in 3D pseudo-strain reconstructions. However, methods are needed for maximising the quality of the data, and ensuring that poor data do not degrade the overall display. Methods in [14] for filtering out bad frames demonstrated scope for identifying good data, without actually achieving a dramatic improvement in terms of the quality of the 3D reconstructions. Bad data are also problematic in 2D image sequences [36]. Similar methods for improved robustness may be useful in both applications.
CHAPTER 1. INTRODUCTION

1.5 Original contributions and thesis outline

This research was aimed at converting the proof-of-principle system from [14] into a clinically valuable system for 2D and 3D elasticity imaging. Progress has been made by developing new signal processing methods and new algorithms for pseudo-strain image formation, while no changes have been made to the scanning hardware. This work is strictly technical, as opposed to medical, but its emphasis was influenced by the opinions of clinical collaborators. Resources and ethical approval for clinical scanning of hospital patients were not available during the period of technical research. Attention was paid to making the system easy to use, while producing reliably meaningful images, so as to support future studies in a clinical setting. Clinical work began during the writing of this thesis, so remarks in the concluding chapter are supported by early clinical examples.

The key to producing good 3D pseudo-strain images is to perform accurate strain estimation where possible, detect the level of accuracy throughout each 2D pseudo-strain image, and construct 3D images based on the best data. This can also be applied to improve live 2D displays. The usual approach mentioned in Section 1.3.1 of recording sequences from which clinicians select good images is impractical and time-consuming, making inefficient use of the available data.

In common with much of the literature on freehand strain imaging, computational efficiency is considered important throughout this work for a decent frame rate (>10 Hz) in the live 2D display, and to avoid long delays in the reconstruction of 3D images. Additionally, three distinctive themes run throughout this thesis:

- **High accuracy.** Fast algorithms are not assumed to be incompatible with highly accurate strain estimation. A range of algorithms — some pre-existing, others novel — are compared to explore the limits of performance.

- **Known accuracy.** The rigorous approach taken in algorithm development supports an extension, so as not only to provide estimates of displacement or strain, but additionally to predict the precision of every estimate.

- **Creative exploitation.** There are many possible ways of exploiting deformation data with high accuracy and known precision. The availability of precision estimates means that pseudo-strain images can be constructed based primarily on the most accurate deformation data.

The research topics divide into two groups. Part II describes work on signal processing methods for producing deformation data from pre- and post-deformation frames of RF ultrasound data. Part III tackles the broader topic of image formation, including filtering that can be applied to deformation data so as to estimate strain, ways of controlling the quality of the results, and superficial aspects of a practical display scheme. Considering the similar approaches anticipated for improving 2D and 3D pseudo-strain imaging, no distinction is made between the two applications until Chapter 6 which describes the practical interface.
Each chapter describes a distinct contribution. A high level of novelty runs throughout this thesis in the context of the pre-existing strain imaging literature, reflecting the unusual emphasis on characterising the level of accuracy in deformation estimates, rather than just trying to improve them. The following brief summary lists the novel contributions in order of significance (in the author’s opinion), the most significant coming last:

- **Dynamic resolution selection** (Chapter 5). Whatever the quality of the data, it is always possible to make strain estimates more accurate by reducing their resolving power. A framework has been devised for estimating strain with a uniform signal-to-noise ratio, regardless of the properties of the ultrasound data. This was a valuable exercise probing the statistical properties of strain estimates, but in hindsight an approach based on radically modulating the resolution may not be entirely desirable.

- **Phase-based estimators** (Chapter 4). A new algorithm for deformation estimation has been developed and tested. It has high performance compared to pre-existing approaches, while offering different opportunities for optimisation, and providing a more versatile basis for future development to further improve performance. This work clarifies some general properties of displacement estimation algorithms, motivated by the following location estimation method.

- **Location estimation** (Chapter 3). For various reasons, a displacement estimate based on pre- and post-deformation ultrasound data may not apply at the expected location. This introduces error in the derivative (strain). Location estimation is shown to be a highly efficient solution. Previous algorithms tackling this problem are far more computationally intensive. Location estimation is probably the most novel concept in this thesis, although not the most significant.

- **Robust tracking** (Chapter 2). Much of the previous literature implies that the most robust framework for deformation estimation is exhaustive search. Tracking methods restrict the search range based on continuity assumptions. This is motivated on the basis of reducing the computational load, but it may also improve accuracy by implicitly invoking prior knowledge. Robust tracking has been approached as a necessary task to support other aspects of this research. In hindsight, details of these heuristic tracking methods are among the more significant contributions of this thesis.

- **Practical 2D and 3D interface** (Chapter 6). The foregoing strands serve as building blocks behind the interface. Pseudo-strain data and precision estimates are processed to modify the superficial properties of the strain imaging system in a way that dramatically increases ease of use and basic accuracy, enabling practical 2D and 3D pseudo-strain imaging by freehand scanning. It is anticipated that this will be clinically useful.
Part II

Deformation estimation
Chapter 2

Robust tracking

2.1 Background

The first signal processing task for generating a strain image is estimation of the deformation throughout the scan region. The inputs are pre- and post-deformation frames of RF ultrasound data. Throughout this thesis, for reasons given in Chapter 1, strain is only estimated in the axial direction. The need to estimate bulk and fine components of axial displacement accurately was mentioned in Section 1.4. Errors in the fine component have previously been described as “jitter”, with errors in the bulk component described as “peak-hopping”; analysis characterising estimation performance has tended to focus on the size of fine errors [143, 144]. This chapter is about methods for correctly estimating the bulk component. Minimising fine errors is irrelevant unless bulk errors are rare. There is non-zero fine error in every displacement estimate, varying in size depending on data quality and algorithm performance, whereas the occurrence of bulk errors is more of a binary event: the local optimum selected as a displacement estimate either is or is not closer to the true displacement than every other local optimum that could have been chosen. The presence of bulk errors can be highly problematic, because they are amplified by later stages of signal processing, impacting disproportionately on strain estimates calculated using linear filters.

All methods for deformation estimation in this thesis share common structural features. An individual estimate of displacement relative to the ultrasound probe is calculated at each point on a grid covering the scan region. A window of data centred on the point of interest in the pre-deformation frame is compared with windows of equal size at shifted positions in the post-deformation frame. A match is identified by calculating the similarity between the windows, noting the post-deformation window that registers the highest similarity. Finally, the displacement estimate is equal to the difference between the positions of the pre- and post-deformation windows. The displacement estimate can have both axial and lateral components if post-deformation windows are tested in multiple neighbouring columns. Otherwise, the lateral component of displacement is treated as zero.

Two features that distinguish between different displacement estimators are the sim-
CHAPTER 2. ROBUST TRACKING

ularity measure and the search range. The correlation coefficient between RF ultrasound data in pre- and post-deformation windows is a popular choice of similarity measure \([13, 135, 136, 145]\) used throughout this chapter. Alternatives are tested in later chapters. Regarding the search range, many authors favour exhaustive search \([13, 135, 136]\), where the search range spans a large fraction or all of the data acquired over the dimension being searched. Correlation coefficients are calculated for every integer shift along the sampled data to find the highest value.\(^1\)

A single RF ultrasound frame spanning a square region of tissue typically consists of around 128 columns of RF ultrasound data, called A-lines, with several thousand RF samples along each A-line. With conventional beamforming the axial dimension in RF ultrasound frames is not only far more densely sampled than the lateral dimension, but it is also less likely to be affected by aliasing.\(^2\) has a higher centre-frequency, and higher bandwidth \([27]\). Therefore, axial displacement estimation is far more accurate than lateral displacement estimation.

Figure 2.1 shows an example of exhaustive search along an A-line. The correlation coefficient is shown for a wide range of trial displacements. The highest peak is near the actual displacement, but wrong peaks occur at intervals roughly equal to the wavelength associated with the centre frequency of the ultrasound signal. The correct peak happens to be the highest in this example, but some of the wrong peaks are almost as high. Bulk errors occur in an exhaustive search whenever the highest peak is wrong.

The rate of bulk errors can be reduced by limiting the search range to include fewer wrong peaks \([145]\). However, bulk errors can only be eliminated if the search range contains the correct peak, and no others — sometimes a neighbouring wrong peak is higher. Ensuring that a limited search range actually includes the correct peak is not trivial. “Tracking” refers to approaches in which displacement is estimated at different points in sequence; each search range is limited based on the values of displacement estimates that have already been calculated at other points nearby. When successful, this eliminates bulk errors.

Tracking in ultrasonic deformation estimation is usually motivated as a means to fast, efficient computation \([33, 80, 138, 146]\). Speed is an important advantage (cf., Section 1.3.1) achieved by shortening the search range, which reduces the number of computations required. Fundamentally this is why exhaustive search is unsuitable for freehand elasticity imaging. However, increased accuracy is also potentially a significant benefit of tracking. In some situations it might even be the main advantage.

Tracking does not necessarily reduce the rate of bulk errors. It causes bulk errors whenever the correct peak lies outside the search range. The first bulk error may skew the search ranges for subsequent displacement estimates, resulting in catastrophic failure.

---

\(^1\)The height of the correct peak may be under-estimated if it falls between integer shifts, so RF ultrasound data at relatively low sample rates are sometimes upsampled (interpolated to increase the number of samples) before an exhaustive search \([69, 136]\).

\(^2\)Anti-aliasing filters applied before digitisation avoid this in the axial direction (along each A-line), but not in the lateral direction (between A-lines).
due to error propagation. On the other hand, strategies for robust tracking that avoid error propagation offer joint benefits of greater accuracy and greater speed.

This chapter describes and compares exhaustive search and various tracking strategies. Displacement estimates are calculated as integer values (number of samples axially and number of A-lines laterally). Results are presented in the form of displacement images, in which bulk errors are easy to identify by visual inspection. Accurate estimation of subsample displacements is considered in later chapters; this is separate to the issue of avoiding bulk errors. Novel strategies introduced here include cross-seeding, multi-pass analysis, continuity checking, secure initialisation, and coarse lateral tracking.

## 2.2 Methods

First, general aspects of the tests are described, including the type of ultrasound data, the choice of similarity measure, and the method for exhaustive search. Second, a range of tracking strategies is described.

### 2.2.1 Tests

Three types of ultrasound data are used for demonstrating differences between the search methods. The width of the scan region is 40 mm in every case. The lateral spacing...
between A-lines is 0.31 mm, and the sample spacing in the axial direction is 0.011 mm (i.e., 910 samples per 1 cm), assuming the speed of sound to be 1540 m s\(^{-1}\). Real ultrasound data were acquired using the system described in Section 1.4. The \textit{in vitro} data are from the scan shown in Figure 1.17, and \textit{in vivo} data are from freehand scanning of a healthy volunteer. Simulated data were generated using Field II [147] with parameters to model the system described in Section 1.4 with a mean SNR of 20 dB from isoechoic scatterer fields before and after uniform-strain deformations.\footnote{Simulations of the same type are used in Chapter 3, where further details are provided.} The pre- and post-deformation RF ultrasound data have a centre frequency of approximately 6.0 MHz, sampled at 66.7 MHz, filtered using a 31-point FIR kernel with a 5–10 MHz passband.

The similarity measure used by way of example is the correlation coefficient between windows in corresponding A-lines, given by:

\[
\rho_{r_1r_2}(n\Delta y, \hat{u}) = \frac{\sum_{n\Delta y} r_1(y)r_2(y + \hat{u})}{\sqrt{\sum_{n\Delta y} r_1(y)^2 \sum_{n\Delta y} r_2(y + \hat{u})^2}},
\]

where \(r_1\) and \(r_2\) are RF data in corresponding pre- and post-deformation A-lines, \(\Delta y\) is the spacing between consecutive windows down the pre-deformation A-line, \(n\Delta y\) is the first sample in window \(n\), \(Y\) is the window length, and \(\hat{u}\) is a trial displacement in the axial direction. Other similarity measures including different correlation coefficients, sum-of-squared-differences or sum-of-absolute-differences [133] could be used interchangeably, without modifying any of the following search methods.\footnote{This is not to say that different similarity measures yield identical accuracy.}

Except where otherwise stated, displacement estimation is performed using windows that are 149 samples long, distributed at intervals of 30 samples along pre-deformation A-lines. The fractional overlap between successive windows in the same A-line is therefore 80\%. \textit{Exhaustive search} is the simplest strategy for displacement estimation. For each pre-deformation window, the similarity measure is calculated with post-deformation windows at a wide range of shifted positions. The displacement estimate is provided by whichever trial displacement registers the highest similarity, as in Figure 2.1. Typically, the set of trial displacements is distributed uniformly over a limited range [13, 135, 136, 145]. A range of ±5 mm (±455 samples) is used in Section 2.3, with one trial displacement per shift of one sample.

### 2.2.2 Tracking strategies

In the following tracking strategies each window is seeded with a first trial displacement, \(\hat{u}'\), and a short-range search finds the nearest peak in the similarity measure, which gives the displacement estimate, \(\hat{u}\). The tracking strategies differ in terms of how seed displacements are selected. In Section 2.3, short-range searches are performed by first calculating three correlation coefficients around the seed: \(\rho_{r_1r_2}(n\Delta y, \hat{u}' - 1)\), \(\rho_{r_1r_2}(n\Delta y, \hat{u}')\) and \(\rho_{r_1r_2}(n\Delta y, \hat{u}' + 1)\). The seed is adopted as \(\hat{u}\) if it has the highest correlation coefficient. Otherwise, further correlation coefficients are calculated at successive integer shifts in the
direction with the higher correlation coefficient. The peak is identified where a further shift reduces the correlation coefficient.

**Basic tracking**

Several authors have used basic tracking in their strain imaging systems [80, 146]. Displacement estimation is performed in each column independently (see Figure 2.2a). The first window is seeded with zero (the relative movement of tissue next to the probe surface is assumed to be small). The second window is seeded with the first displacement estimate, the third window with the second estimate, and so on (the relative displacement between windows at consecutive depths is assumed to be small). The bulk component of displacement accumulates along each column.

**Cross-seeding**

The limitations of basic tracking are exposed easily. Cross-seeding is a novel strategy offering greater robustness. Columns are processed in parallel, one row at a time. The seed for a particular window is the displacement estimate from the previous row with the highest associated correlation coefficient (or other similarity measure) within \(L\) columns to either side. A high correlation coefficient does not guarantee that an estimate is correct, but it does indicate relatively low likelihood of bulk error. The rate of bulk errors can be reduced by choosing a suitable value for \(L \geq 1\), so that information can flow laterally across the image (see Figure 2.2b).

**Multi-pass analysis**

Multi-pass analysis further reduces the error-rate. The first pass runs from the top to the bottom of the image, seeding each window from the row above. For the second pass, the direction is reversed, running from the bottom to the top of the image, seeding from below. A first displacement estimate and associated correlation coefficient are calculated at every window during the first pass, and stored in a display buffer. During the second pass, new estimates are calculated when the seed differs from the existing value in the display buffer by more than a threshold (two samples is used in Section 2.3). When a new estimate is calculated, it enters the display buffer if the new correlation coefficient is higher. Estimates with lower correlation coefficients are stored in a reserve buffer.

**Continuity checking**

There is a risk with multi-pass analysis that some correct displacement estimates may enter the reserve buffer owing to low correlation coefficients. Nevertheless, each value in the display buffer is more likely to be correct than the corresponding reserve value. Continuity checking is a low-cost heuristic for correcting the small fraction of estimates that

\[\text{†}\]A short-range search near the existing displacement estimate is redundant, because it converges on the same value.
are sorted incorrectly. Working through each window in turn, the average displacement is calculated of the four neighbouring values in the display-buffer. The value from the reserve buffer swaps into the display buffer if it offers greater continuity (i.e., is closer to the average, see Figure 2.2d).

**Secure initialisation**

The premise behind cross-seeding and multi-pass analysis is that correct estimation propagates preferentially throughout each image whenever some good displacement estimates are produced. This can fail if no single good estimate is found from which to propagate the correction. Trivially, a pair of pre- and post-deformation ultrasound frames might be so different as to make displacement estimation impossible. For example, the probe may move so far elevationally that there is no overlapping data (see Figure 1.16). However, it is also possible that no good displacement estimates are produced, despite the pre- and post-deformation ultrasound frames being sufficiently correlated, in which case the initialisation has failed.

Basic tracking is initialised by seeding windows at the probe-face with a trial displacement of zero. In real ultrasound images there is often a layer at the top dominated by reverberation associated with the coupling between the probe and the tissue, which is stationary relative to the probe. However, near-zero displacement is not necessarily representative of the top-most echos from the tissue, because the tissue surface can move a couple of millimetres up or down under a thick layer of gel or other coupling fluid. Initialisation with zero at the top of the image can also fail if the top layer of tissue is fairly anechoic and very soft.

**Secure initialisation** can significantly improve overall robustness. The initialisation is less likely to fail if it is performed deeper within the ROI, avoiding artefacts associated with the contact layer. Examples in Section 2.3 are initialised 30% of the way down the image. Short-range searches at this depth obviously cannot be initialised with zero. Two approaches have been devised allowing for the unknown displacement at the initialisation depth.

Exhaustive search can be used for displacement estimation in the initialisation row. Correct estimation propagates into the rest of the image by cross-seeding and multi-pass analysis, provided that at least one of the exhaustive searches yields a correct estimate with a fairly high correlation coefficient.

Alternatively, windows in the initialisation row can be seeded with a wide range of trial displacements, many of which are wrong. This can be implemented as a saw-tooth pattern of seed displacements along the initialisation row. Correct displacement estimation propagates from whichever windows happen to have been seeded correctly. The saw-tooth used in Section 2.3 spans $\pm 2.2$ mm.

Whichever version of secure initialisation is used (exhaustive or saw-tooth), it needs to be combined with cross-seeding, and multi-pass analysis including continuity checking. Tracking propagates down from the initialisation row to the bottom of the image, then back from the bottom to the top, and finally returns down to the initialisation depth, so that every window is visited twice (see Figure 2.2e).
CHAPTER 2. ROBUST TRACKING

Figure 2.2: Tracking strategies: (a) Basic tracking. Each column is processed independently. The top window is seeded with zero displacement. Each lower window is seeded with the estimate from the window above. Large displacements accumulate along the columns. (b) Cross-seeding. Each window is seeded with a high-correlation estimate from the previous row within \( L \) columns to either side. The subscripts indicate the coordinates of each window and the numbers in brackets are correlation coefficients. Seeds are shown propagating through four rows in three neighbouring columns with \( L = 1 \). (c) Multi-pass analysis. Tracking with cross-seeding proceeds first from the top to the bottom of the image, and second from the bottom to the top. (d) Continuity checking. Some windows have two displacement estimates after multi-pass analysis. Estimates with higher correlation coefficients enter the display buffer, while estimates with lower correlation coefficients are stored in a reserve buffer. Continuity checking retrieves correct estimates that have been forced erroneously into the reserve buffer. (e) Secure initialisation. Tracking is initialised at a row deeper within the ROI, either by exhaustive search or with a saw-tooth pattern. Multi-pass analysis proceeds to the bottom, back to the top, and finally down to the initialisation depth.

Coarse lateral tracking

Multi-dimensional tracking is important if substantial multi-dimensional motion occurs, otherwise data misalignment causes axial displacement estimation to fail. The non-axial motion is often small enough that it can be ignored. When this is not the case, there is unfortunately no way to track elevational movement using ultrasound data acquired in
a 2D format, but there is scope for coarse lateral tracking to avoid failure due to lateral movement.

Coarse lateral tracking is an efficient strategy for data alignment so that pre-deformation windows are compared with the correct post-deformation A-line in the presence of large lateral motion. A suitable method is adopted from Treece et al. [148]. The initialisation row begins with zero lateral displacement. Experience to date has not indicated a need for more sophisticated initialisation. At the end of each short-range axial search, two extra correlation coefficients are calculated for post-deformation windows in the neighbouring A-lines with the same axial displacement. The lateral displacement estimate of the current window is incremented if a higher correlation coefficient is found in either neighbour. Multi-A-line lateral displacement estimates accumulate over the course of cross-seeding and multi-pass analysis. Although coarse lateral displacement estimates are useful for alignment purposes, they have no intrinsic value; they are not suitable for calculating lateral strain.

2.3 Results

Exhaustive search and tracking strategies are demonstrated in a series of examples. These provide qualitative illustrations of situations in which bulk errors occur, and show how different strategies avoid them. A tracking strategy either succeeds or fails depending on qualitative features of the tissue motion. Each strategy has different failure modes. The advantage of the more robust strategies is a reduction in the range of failure modes, without unacceptably increasing the computational cost.

**Basic tracking versus exhaustive search**

Figure 2.3 shows a simulation, from which Figure 2.3a is a typical B-scan. The data properties are fairly uniform, although the focusing and SNR are poorer towards the top and bottom of the image. There are no regions of especially bad data where large errors would be inevitable. Displacement images for a 1% compressive strain are shown in Figure 2.3b on a scale of black–white representing displacements of 0–35 samples towards the probe.† Material towards the bottom of each image has moved furthest towards the probe.

Exhaustive search and basic tracking are compared using two different window lengths.‡ Note that the ultrasonic wavelength at the probe centre frequency corresponds to approximately 11 samples, so the lengths can optionally be thought of in number of wavelengths. Exhaustive searching with a window length of 61 samples mostly finds the correct displacement in Figure 2.3b1, but there are a significant number of outliers. Basic tracking with the same window length is more successful in Figure 2.3b2 — there are no bulk

---

†In all of the images, displacements off the scale are clipped to the display range.

‡The windows have a width of 1 A-line. 2D windows can be used, spanning multiple A-lines. This does not significantly change the relative merits of different tracking strategies.
Figure 2.3: Simulation with uniform strain. (a) Typical B-scan. (b) Displacement images with 1% strain. (b1) 61-sample windows, exhaustive search. (b2) 61-sample windows, basic tracking. (b3) 149-sample windows, exhaustive search. (b4) 149-sample windows, basic tracking. (c) Displacement images with 4% strain. (c1) 149-sample windows, exhaustive search. (c2) 149-sample windows, basic tracking. (c3) 149-sample windows, cross-seeding, $L = 128$. (c4) 149-sample windows, cross-seeding, $L = 1$. (N.B. All ultrasound images in this thesis represent a physical width of 40 mm, and are displayed in the correct aspect ratio.)
errors. Increasing the window length to 149 samples in Figure 2.3b3 almost eliminates the bulk errors in exhaustive search. Basic tracking in Figure 2.3b4 is again free of bulk errors.

Windows of 149 samples are slightly longer than ideal for displacement estimation in a simulation of 4% strain. The scale in Figure 2.3c is black–white representing displacements of 0–139 samples towards the probe. Exhaustive search in Figure 2.3c1 is disastrously unsuccessful. Basic tracking is still free of bulk errors throughout most of Figure 2.3c2, but where they arise errors propagate to fill the column. Basic tracking is better than exhaustive search until the first error, but thereafter it fails completely.

Cross-seeding

Cross-seeding in Figure 2.3c3–4 produces much better results than exhaustive search or basic tracking. These examples are the two extremes of cross-seeding. Setting $L = 128$ means that every window in a row is seeded with the same displacement estimate of highest correlation coefficient from the previous row. In principle, there is a risk of the highest correlation coefficient being misleading, which could result in tracking failure throughout the image below a certain row. However, the highest correlation coefficient usually indicates a particularly reliable displacement estimate, as in Figure 2.3c3. Obvious displacement estimation errors are restricted to a few small dots, which are just visible on close inspection. Figure 2.3c4 with $L = 1$ contains a few small regions where a wrong seed is taken from the current column or a direct neighbour, leading to small patches that may be described as bulk errors, since Figure 2.3c3 shows that clearly a more-correct correlation peak exists. In this case there is no risk of the image spontaneously going wrong below a certain row, but there is a higher probability of small error regions emerging, where a bulk error propagates from one row to the next.

Related work has been undertaken concurrently yet independently by Zahiri-Azar et al. [149]. Their method is quite different from cross-seeding, although they do use correlation coefficients to assess reliability. Basic tracking is used by default, replaced by an exhaustive search whenever the correlation coefficient drops below a threshold of 0.9. This avoids dropouts due to tracking failure, but compared to cross-seeding it has less of a speed advantage, and ultimately less potential for improved accuracy over exhaustive search. In the example of Figure 2.3c, the method of Zahiri-Azar et al. would revert to exhaustive search for the entire image. The majority of displacement estimates throughout Figures 2.4–2.7 would also end up being calculated by exhaustive search.

Returning to cross-seeding, the success in Figure 2.3c3 with $L = 128$ owes to a special property of the simulation: the displacement does not vary with lateral position. This is why a single seed can be applied successfully across an entire row. Real scan data from a non-ideal target better illustrate practically significant differences between different levels of cross-seeding. In particular, tracking is difficult principally because the quality of real data tends to be inhomogeneous. Figure 2.4 is an example of real ultrasound data,
where the displacement is mostly easy to track. Black–white in the displacement images represents 0–14 samples displacement away from the probe. Bulk errors from exhaustive search appear in Figure 2.4b at boundaries of the olive and in the region of high strain at the top. However, basic tracking in Figure 2.4c actually does worse, with two columns of bulk errors, starting from points on the upper surface of the olive where the displacement estimates are noisy. This shows why cross-seeding is needed. However, cross-seeding with $L = 128$ is unsuccessful in Figure 2.4d, because displacement varies substantially within individual rows. In this example, seeding from neighbouring columns ($L = 1$) is the most successful approach — Figure 2.4e appears to be free of bulk errors. The rate of variation in displacement over the lateral direction is finite, so a compromise of $L = 10$ does nearly as well in Figure 2.4f, apart from a couple of errors on the upper surface of the olive.

**Multi-pass analysis and continuity checking**

When exhaustive search is used for displacement estimation in the example of Figure 2.5 it produces a large number of bulk errors, but is nonetheless more successful than the tracking methods tested so far. The scale of displacement images in Figure 2.5 is 0–15 samples towards the probe. Basic tracking in Figure 2.5c fails across almost the entire image. The reason is indicated by the cross-seeding results in Figure 2.5d1 and e1. The
displacement estimates at the top of the image are seeded with zero, which leads to errors right across the top row except at the right-hand end. Evidently the probe rotated while this frame was acquired, so the surface of the gelatin moved substantially away from the probe, except on the right-hand side where it is in direct contact. With cross-seeding the initially small number of correct displacement estimates on the right propagates diagonally into the rest of the image. However, half of the image remains filled with errors for $L = 1$ in Figure 2.5d1, or a smaller fraction for $L = 10$ in Figure 2.5e1 where the correction
A second pass up the image enables algorithms with cross-seeding to fill (Figure 2.5e2) or almost fill (Figure 2.5d2) the entire frame with correct displacement estimates. Note that the second pass involves many fewer computations than the first, because most seeds differ from the estimate of the first pass by less than the recalculation threshold. New estimates are recalculated in Figure 2.5d2 and e2 almost exclusively in the parts that change from Figure 2.5d1 and e1. For many images, the second pass merely consists of checking that the seeds are similar to existing displacement estimates, without performing any recalculation. This is an effective and inexpensive way of enabling correct displacement estimation to propagate throughout the scan region, although it is not perfect. The correction propagates slowly in Figure 2.5d with \( L = 1 \), leaving a triangle of incorrect displacement estimates even after the second pass. That triangle could obviously be corrected by continuing with a third pass. In principle, passes could repeat until they meet a stopping criterion (e.g., no value changing in either the display or reserve buffer). However, the simplicity of a two-pass approach is appealing. It almost never arises that three or more passes increase the accuracy beyond that achieved by two-pass analysis when the lateral propagation of correct seeds is fast (e.g., \( L = 10 \)).

Within regions that are successfully tracked, a few bulk errors remain after the second pass. The region of low signal, below the olive, appears noisy in both Figures 2.5d2 and e2, but these estimates may represent the correlation coefficient peaks that are nearest to the true displacement. More significant failures are the obvious outliers towards the top right of the image. These are points where incorrect displacement estimates have high correlation coefficients, which should be dealt with by continuity checking. This replaces all of the outliers with good estimates in Figure 2.5e3. One outlier remains in Figure 2.5d3, at a point where the reserve buffer was empty.

**Secure initialisation**

The difficulty with tracking in the previous example is partly the result of failed initialisation. Figure 2.6 shows a more extreme example. The scale of the displacement images is 0–25 samples away from the probe. Again, exhaustive search indicates the rough form of the displacement field, but it includes many outliers. The upper surface displaces substantially relative to the probe, right across the image, so displacement tracking with cross-seeding fails if it starts with zero from the top of the image (see Figure 2.6c).

Much greater success is achieved by the tracking methods with secure initialisation (Figure 2.6d–e). The exhaustive initialisation in Figure 2.6d throws up a few outliers alongside the good seeds, but they fail to propagate. Figure 2.6d2 shows the outcome after every window has been visited twice in multi-pass analysis. The disadvantage of exhaustive initialisation is that outliers at the initialisation depth have high correlation coefficients, so only continuity checking can remove them from the display buffer. The saw-tooth initialisation also performs well in Figure 2.6e. The two forms of secure initialisation lead to identical images after continuity checking in Figure 2.6d3 and e3. It is difficult to judge which approach is likely to be more robust in general. What is cer-
Figure 2.6: Displacement images from the scan of an olive in gelatin (very difficult example). (a) B-scan. (b) Exhaustive search. (c) Cross-seeding, $L = 10$. (d) Cross-seeding, $L = 10$, with secure initialisation by exhaustive search, (d1) after pass 1a, (d2) after pass 2b, (d3) after continuity checking. (e) Cross-seeding, $L = 10$, with secure initialisation by the saw-tooth method, (e1) after pass 1a, (e2) after pass 2b, (e3) after continuity checking.

One of the images presented up to this point can be improved by lateral tracking, but substantial lateral movement can occur. Figure 2.7 shows examples in which coarse lateral

Coarse lateral tracking

None of the images presented up to this point can be improved by lateral tracking, but substantial lateral movement can occur. Figure 2.7 shows examples in which coarse lateral
Figure 2.7: Displacement images from scans with lateral movement. All examples include cross-seeding ($L = 10$), multi-pass analysis, continuity checking and the saw-tooth form of secure initialisation. (a) Simulation of 2% axial strain, with 1% lateral strain, as per a Poisson’s ratio of 0.5, and further lateral motion equal to 2% of depth due to shearing and rotation. (a1) Axial displacement estimates without coarse lateral tracking. (a2) Axial displacement estimates with coarse lateral tracking. (a3) Lateral displacement estimates produced by coarse lateral tracking. (b) Frame with noticeable side-slip from a freehand scan of normal breast tissue. (b1) Axial displacement estimates without coarse lateral tracking. (b2) Axial displacement estimates with coarse lateral tracking. (b3) Lateral displacement estimates produced by coarse lateral tracking.

tracking is helpful. The simulation in Figure 2.7a verifies that the strategy described for coarse lateral tracking behaves reasonably. In addition to a uniform axial strain of 2%, the scatterer field has undergone lateral strain of 1%, as per incompressibility, and a combination of shear and rotation amounting to movement to the right varying linearly with depth (lateral shift equal to 2% of the depth). This means that the maximum displacement in the bottom right of the image is 0.8 mm towards the probe (axially) and 1.0 mm to the right (laterally).

The scale of the axial displacement images in Figure 2.7a1–2 is 0–70 samples towards the probe. Without coarse lateral tracking, bulk errors cover a large fraction of Figure 2.7a. Some axial displacements are tracked successfully, even where the lateral displacement is similar to the spacing between A-lines. The poor lateral resolution of the ultrasound means that lateral movement up to around 0.3 mm leaves sufficiently corre-
lated data in the original A-line such that bulk errors are not inevitable. However, the axial displacement estimates are clearly significantly more accurate where coarse lateral tracking is used in Figure 2.7a2. It is interesting to inspect the values of the lateral displacement estimates. There are five levels of lateral displacement in Figure 2.7a3, where black–white represents -1–3 A-lines to the right. The bands appear exactly where they should for this simulated displacement field. However, while this tracking algorithm was not designed to maximise the accuracy of the lateral displacement estimates, the images are indicative of the far lower potential of conventionally-beamformed ultrasound data for supporting accurate displacement estimation in the lateral direction.

The practical benefit of coarse lateral tracking is indicated by Figure 2.7b. These are tracking results for in vivo freehand scanning. The axial displacement estimates (shown in Figure 2.7b2 on a scale of 0–20 samples away from the probe) are much more convincing with coarse lateral tracking. The values of the lateral displacement estimates are shown in Figure 2.7b3 (black, grey and white represent displacement of 0, 1 and 2 A-lines to the right). Evidently the lateral movement was mainly caused by the probe slipping to the left, since the improvement in Figure 2.7b2 is achieved mostly by tracking displacement at a uniform lateral shift of 1 A-line.

2.4 Discussion

The key strategy for robust tracking is cross-seeding. A correct displacement estimate is used to seed the search at a nearby window. Bulk displacement information propagates from one window to the next. Problems only arise if the bulk displacement estimate is thrown off course by a large error subsequently used as a seed. Cross-seeding exploits the fact that the absolute value of a similarity measure such as correlation coefficient indicates the likelihood of bulk error. Correct bulk displacement information steers around patches of bad ultrasound data on a path determined by the correlation coefficients. Cross-seeding with $L = 10$ is an effective way of achieving this, while allowing for variation in the displacement across each row.

Another crucial aspect of robust tracking is secure initialisation to ensure that some correct estimates are found from which to propagate the correction. Initialising deeper within the ROI is more robust for many applications, since coupling artefacts are common near the top of ultrasound images [27].

Multi-pass analysis is probably the least general of the strategies described. It allows correct bulk displacement information to propagate further than it can in a single pass. Single-pass tracking is highly directional; information only flows down the image or diagonally downwards. The overall robustness is increased when information can be passed by seeds between any pair of windows in any direction. For cross-seeding in the form described here, multi-pass analysis is required so as to enable bulk displacement information to flow in arbitrary directions. This was necessary in Figure 2.5 owing to an initialisation problem. There are many other situations where it also helps. Tracking usually fails in large blood vessels, because data within the vessel are uncorrelated from one ultrasound
frame to the next. When a blood vessel is scanned transversely, tracking can recover by propagating around the side. However, a patch of errors develops underneath the vessel unless \( L \) exceeds its radius, and the patch is only corrected by a second pass.

A more optimal form of cross-seeding may avoid the need for multi-pass analysis. A more elegant scheme for achieving direction-independence has been demonstrated [150], extending the strategy described in this chapter. However, it bears emphasising that multi-pass analysis is not particularly cumbersome, being on average only slightly more costly than single-pass analysis.

Continuity checking may actually give multi-pass analysis an advantage over single-pass non-directional algorithms such as [150], because multi-pass analysis often provides two displacement estimates where the data are of lowest quality. Continuity checking to sort between the displacement estimates is effectively a nonlinear filter that not only detects outliers, but can also replace them with independent, correct estimates. Whether features like this offer measurable performance improvements compared to non-directional cross-seeding has not been demonstrated, but this issue may be worth investigating in further research.

The other strategy described is coarse lateral tracking. In a sense, this is a general estimation issue rather than a tracking feature. It is useful to note that lateral motion can be followed by robust tracking, and this does prevent bulk errors due to lateral misalignment. However, the significance of lateral tracking is application-dependent, even though all deformed tissues exhibit both axial and lateral strain. If the probe face is flat, the main strain is likely to be axial. High strains occur in small regions, but freehand scanning rarely involves axial strain larger than 1% over a large part of an image. Suppose that an entire image consisting of 128 A-lines is subject to 1% axial strain. This is likely to be accompanied by 0.5% lateral strain, so tissue at the edge of the image moves 0.3 A-lines laterally. This is below the threshold at which coarse lateral tracking could be useful. Significant lateral motion is more likely when the probe slips on the surface, or organs translate along slip planes within the body. Lateral tracking is also useful in conjunction with convex (curved) probes, where large motion perpendicular to the axis of the ultrasound beam is more likely.

**Statistics**

Tracking methods can always be motivated by arguments relating to computation time [33, 80, 138]. However, robust tracking strategies may produce displacement estimates with lower rates of bulk error than alternatives including exhaustive search. This is significant in terms of the uses to which the data can then be put. Characterisation of displacement estimation errors has been an active research topic for decades, primarily motivated by military applications based on sonar and radar [151, 152]. Theory for characterising the size of errors is relevant to quasistatic elasticity imaging. The usual approach is to evaluate a Cramér-Rao lower bound on the estimation variance [81, 143, 144, 153, 154], which is effectively an estimate of the mean squared error when bulk errors do not occur. Theory allowing for the possibility of bulk errors leads to much higher lower bounds on
estimation variance [81, 155, 156, 157]. The wide discrepancy arises due to the fairly small ratio of bandwidth to centre frequency (usually < 1.0) of typical RF ultrasound signals [155, 158, 159].

When bulk errors are avoided, the estimation variance is not only smaller, but also relatively predictable. Tangible quantities such as the height of the peak in correlation coefficient are related to the extent to which noise can introduce error in the position of the peak. This is not the case when bulk errors arise, in which case the scale of errors is almost entirely unpredictable. From a statistical perspective, robust tracking can be thought of as guiding the displacement estimator towards the Cramér-Rao lower bound, making the displacement data more manageable for subsequent processing such as the techniques described in Chapters 5 and 6.

Alternatives

It would be wrong to suggest that robust tracking strategies are the only means by which to improve the statistical properties of displacement estimates. Nonlinear filtering for outlier removal can perhaps be devised to clean up the final output from exhaustive search. On the other hand, if substantial numbers of outliers are removed, nonlinear filtering cannot replace them with independent non-outlier displacement estimates. This will often be an advantage of robust tracking, although it is unlikely that robust tracking can eliminate bulk errors entirely, so nonlinear filtering remains a worthwhile topic for further research.

In terms of search methods, robust tracking is still partly motivated by computational efficiency. Greater robustness could be achieved if computational cost were not an issue. For example, every peak in correlation coefficient for every window in an ultrasound frame could be calculated and stored. Correlation coefficients could be converted into relative probabilities associated with each peak, and the plausibility of different levels of displacement continuity could be modelled with a prior distribution. A relative \textit{a posteriori} probability could be calculated for every possible combination of displacement estimates across all of the windows. The maximum \textit{a posteriori} set of displacement estimates would presumably be more robust than any form of tracking. However, the combinatorial number of displacement fields to be tested would rise exponentially with the number of windows. Naively computing every probability value would rapidly become infeasible regardless of time and computing power. An efficient search method will always be necessary. Robust tracking is an attempt at devising a highly efficient search strategy that implicitly reflects prior assumptions regarding displacement continuity.

Another search method aiming for robustness is multi-scale displacement estimation [160, 161, 162]. The displacement estimates again reflect implicit continuity assumptions, but the failure modes are different. Multi-scale searches fail especially dramatically if errors occur at an early, coarse stage of the search [160, 161]. In any event, multi-scale search methods are much more computationally expensive than robust tracking [160, 161, 162]. Different types of signal are required for different stages of multi-scale estimation, because the phase of RF ultrasound data introduces confusion when searching
CHAPTER 2. ROBUST TRACKING

on a coarse scale. Envelope data is used for the early stages of multi-scale searches [160, 161, 162], but the full signal (including phase) is required for fine-scale estimation to maximise the level of accuracy [162].

Phase-based estimators

Tracking is a prerequisite for the use of phase-based displacement estimators. Displacement can be estimated by aligning pre- and post-deformation ultrasound signals so that their phases match. This has advantages discussed in the following chapters, but the bulk component of displacement must be estimated by some other means. Robust tracking provides a solution so long as a similarity measure such as correlation coefficient or computationally cheaper alternatives is calculated once for every displacement estimate found by phase-based searching. A phase-based estimator was used in conjunction with basic tracking in the preliminary investigation of freehand 3D elasticity imaging, but the vulnerability of basic tracking was found to be a severe disadvantage [14].

2.5 Conclusion

The development of tracking strategies has never been the main research focus during this PhD. Novel tracking strategies emerged over time to support the other strands of technical development. In fact, robust tracking is the foundation of the signal processing methods described throughout the rest of this thesis. The techniques all work less well in conjunction with exhaustive search or basic tracking, because it is vital to avoid bulk errors. The simple, heuristic strategies introduced in this chapter are easy to implement, and may be of use to other engineers working with similar problems.

However, the current set of strategies must not be regarded as definitive. This is fertile territory for further research to further reduce the range of failure modes. For example, no method has been provided for displacement estimation in multiple continuous regions separated by boundaries of discontinuous displacement. Tracking with cross-seeding, multipass analysis and secure initialisation often successfully estimates displacement on both sides of a slight discontinuity, such as a slip plane, but it can fail in extreme cases. Chen et al. have extended these strategies to successfully estimate displacement above and below a pulsating artery viewed longitudinally [150]. The strategies described in this chapter fail in that scenario.

Robust tracking seems to be a fast and effective way of avoiding bulk errors. The literature on tracking methods is surprisingly small considering this observation, which may partly reflect the fact that improvements in tracking methods are difficult to quantify. By contrast, methods for minimising fine-scale displacement estimation errors are a more popular topic for scientific articles. Progress with fine-scale estimation can be quantified easily in terms of the estimation variance [81] as in Chapters 3 and 4, but pragmatic research to improve tracking and related techniques is just as important for efficient, robust deformation estimation.
Chapter 3

Location estimation

3.1 Background

Although Cramér-Rao lower bound theory is popular for characterising fine errors in deformation estimation by window-matching [81, 143, 144, 154, 163, 164, 165, 166, 167, 168], such analysis is only strictly applicable when the displacement is stationary [81]. In other words, a uniform displacement should apply throughout each window of ultrasound data, as in rigid-body motion. Ironically, only the non-rigid component of displacement is of interest in elasticity imaging, where strains are calculated by differentiation (cf., Section 1.2). Each window still provides an estimate of the average or typical displacement over the region that it covers, but there is ambiguity regarding the location at which each displacement estimate applies with greatest validity. This is a source of error when displacements are differentiated to calculate strain. In effect, the estimated deformation comprises a set of displacement/location data-pairs, in which location errors are potentially significant. This chapter analyses and tests several methods for avoiding location error.

The assumption that displacement estimates apply with greatest validity at the centre of each window gives rise to location error, because it is not generally correct. Céspedes and Ophir [169] conducted an early study of window-matching by correlation coefficient methods, noting that the actual location is skewed towards sections where the windowed ultrasound signal has high amplitude. This causes artefacts at boundaries between regions of differing echogenicity, as shown in Figure 3.1, where the strain estimates are corrupted by unwanted amplitude modulation (AM). This can lead to severe misinterpretation of strain images, because AM artefacts correlate with features in B-scans. AM also degrades strain estimates within regions that are isoechoic, where the amplitude still varies due to speckle formation [27].

The size of location errors can most simply be limited by using short windows [145], but this is accompanied by reduced accuracy in the values of the displacement estimates [154]. It might be that low-accuracy displacement estimates from short windows can be used in combination with sophisticated post-filtering to produce good strain images [170, 171], but this chapter investigates the scope for producing accurate displacement/location data-
Figure 3.1: (a) B-scan of RF data from a scan of human forearm. The signal is compressed to simulate a uniform compressive strain of 1% — this is performed by interpolating and resampling the data. On a linear scale from black (0% strain) to white (2%), this should produce a uniform strain image with extremely low estimation noise, since signal noise and decorrelation are lower than could possibly be achieved in a real compression scan. However, (b) phase zero estimation [80] produces a strain image that is severely degraded (and misleading) owing to amplitude modulation (AM), while (c) shows the near perfect result when the same algorithm is enhanced by the best of the correction techniques introduced in this chapter. For both images, displacement estimation windows were $13.5\lambda$ in length, where $\lambda$ denotes the ultrasonic wavelength at the centre frequency.

Pairs using windows of finite length, since this should produce more accurate displacement data when AM is handled appropriately. Lateral and elevational displacement estimates are likely to be affected by AM when a multi-dimensional approach is taken [135], and location errors potentially reduce the accuracy of any estimated strain components, e.g., shear strain [172]. However, in common with the rest of this thesis, this chapter focuses on axial strain estimation.

Céspedes and Ophir suggested two techniques for AM suppression [169]. Firstly, compression of the signal amplitude reduces fluctuations, thereby shifting the most-valid location towards the centre of each window. This is an effective means of AM suppression, which has been reapplied in more recent strain imaging systems [14, 80]. Alternatively, adaptive stretching compensates for nonuniform displacement by applying a uniform stretch to the windowed signal, so potentially the displacement estimate is valid at all points in the window, provided that the within-window strain is uniform [169]. Numerous studies have demonstrated that adaptive stretching reduces strain estimation noise, although it incurs significant computational cost [136, 137, 173, 174, 175].

The following section examines AM from a theoretical standpoint, leading to surprisingly simple AM correction (AMC) by location estimation. Experiments have been performed using simulated RF ultrasound data to compare phase zero and correlation coefficient methods for deformation estimation, and to evaluate the efficacy of AMC compared
to amplitude compression. Both methods are computationally efficient, hence suitable for real-time imaging systems. Further experiments have been performed using adaptive stretching, which is slower, but provides an AM suppression benchmark.

### 3.2 Theory

Theoretical analysis of the estimation noise due to AM leads to AMC, which is a general correction method potentially applicable to any displacement estimator based on window-matching. Two examples of AMC are presented, and pre-existing AM suppression methods — amplitude compression and adaptive stretching — are discussed in the light of this analysis. All of these methods are tested empirically in Sections 3.3 and 3.4.

#### 3.2.1 AM noise

The extent to which location error translates through to strain error depends on the filter used for strain estimation, but relatively sophisticated filters (see Chapters 5 and 6) do not alter the fundamental principle of AM. For the sake of clarity, this analysis refers simply to estimating axial strain, $\hat{\varepsilon}$, by differencing successive axial displacement estimates, and dividing by the spacing between them:

$$\hat{\varepsilon} = \frac{\hat{u}_2 - \hat{u}_1}{\hat{y}_2 - \hat{y}_1},$$

(3.1)

where $\hat{u}_1$ and $\hat{u}_2$ are displacement estimates from windows 1 and 2 respectively, and $\hat{y}_1$ and $\hat{y}_2$ record the estimation locations. It will be shown that these are not generally the same as the centres of the windows. It is sometimes assumed that Equation 3.1 contains only two random variables, $\hat{u}_1$ and $\hat{u}_2$ [81], but this chapter examines the other variables, $\hat{y}_1$ and $\hat{y}_2$. Symbols $\hat{D}$ and $\hat{F}$ are defined as:

$$\hat{D} = \hat{u}_2 - \hat{u}_1,$$

(3.2a)

$$\hat{F} = \frac{1}{\hat{y}_2 - \hat{y}_1},$$

(3.2b)

$$\hat{\varepsilon} = \hat{D}\hat{F}.$$  

(3.2c)

The two sources of estimation noise are illustrated in Figure 3.2. Errors in $\hat{D}$ and $\hat{F}$ are expected to be uncorrelated and unbiased\[^1\] so strain estimation variance, $\sigma^2_{\varepsilon}$, can be expressed in a simple form:

$$\sigma^2_{\varepsilon} = \sigma^2_D\sigma^2_F + \mu^2_F\sigma^2_D + \mu^2_D\sigma^2_F,$$

(3.3)

\[^1\]Uniform bias in $\hat{u}_n$ or $\hat{y}_n$ is removed by the differencing operations that produce $\hat{D}$ and $\hat{F}$. If the errors are nevertheless correlated, the direction of error in $\hat{D}$ must be related to the direction of error in $\hat{F}$. It is hard to envisage how this could ever arise. Error in $\hat{F}$ is related to the degree of correlation between successive errors in $\hat{u}_n$, but it does not indicate the sign of the difference between successive errors (which amounts to the direction of error in $\hat{D}$).
Figure 3.2: A practical displacement estimate is shown between two ideal estimates, where estimation error causes it to deviate from the actual displacement in the underlying tissue. The labelled arrows indicate the two sources of noise.

where $\mu$ denotes statistical expectation and $\sigma^2$ denotes variance. Subscripts $\hat{D}$ and $\hat{F}$ indicate displacement and location factors respectively. The term “variance” is used synonymously with “mean squared error”, so a fixed value of $\hat{F}$ only implies zero variance if $F$ is indeed constant.

What impact do the terms in Equation 3.3 have on strain image quality? An indicator referred to as estimation signal-to-noise ratio, $\text{SNR}_e$ [81, 169], can be measured experimentally in images where the underlying strain field is known to be homogeneous:

$$\text{SNR}_e = \frac{\mu_{\hat{E}}}{\sigma_{\hat{E}}}.$$  \hspace{1cm} (3.4)

$\mu_{\hat{E}}$ is either the actual strain or the mean strain estimate in a homogeneous region, and $\sigma_{\hat{E}}$ is the root mean squared error. The presence of $\mu_D^2$ in the third term of Equation 3.3 is important. While any increase in the strain signal, $\mu_{\hat{E}}$, reduces the significance of displacement errors, the location errors caused by AM are effectively scaled by the strain, so AM can become a dominant source of noise at high strains:

$$\text{SNR}_e = \left( \frac{\sigma_D^2 \sigma_{\hat{F}}^2 + \sigma_D^2/\Delta y^2}{\varepsilon^2} + \sigma_{\hat{F}}^2 \Delta y^2 \right)^{-\frac{1}{2}}.$$  \hspace{1cm} (3.5)

Equation 3.5 comes from substituting the right-hand side of Equation 3.3 into Equation 3.4, with $\Delta y$ denoting the spacing between windows. This result incorporates three simplifying assumptions. (a) $\mu_{\hat{E}} = \varepsilon$, i.e., the mean strain estimate is equal to the true strain. For example, if the mean strain estimate does not equal the true strain, then the displacement estimates must become increasingly biased as a function of distance, which does not occur in practice. (b) $\mu_{\hat{F}} = 1/\Delta y$, i.e., the mean estimate of the reciprocal location spacing equals the reciprocal of $\Delta y$ \(^{1}\). (c) $\mu_D = \varepsilon \Delta y$, i.e., the mean displacement between neighbouring windows equals the strain multiplied by the window spacing \(^2\).

\(^{1}\)This is precisely true when means are calculated by averaging over image area, which is the approach taken throughout this work.

\(^{2}\)This is an approximation, since the mean location spacing averaged over image area is slightly larger than $\Delta y$. In any event, $\mu_D$ is proportional to $s \Delta y$, and none of the following arguments changes if the constant of proportionality is slightly greater than 1.0.
Equation 3.5 indicates the significance of the strain level. Note that $\sigma_D^2$ can usually be reduced by increasing the window length, $Y$ [154]. Reasonable assumptions suggest a relationship of direct proportion between window length and SNR$_e$ when AM is negligible (see Appendix A.1), so the occurrence of such a relation would indicate successful AM mitigation. The range of possible values of $F$ increases with $Y$, however, so AM noise in $\sigma_F^2$ could make it difficult to exploit long windows.

Of course, none of this matters if practical displacement estimators really produce estimates with greatest validity at window centres, i.e., if $\sigma_F^2 = 0$, in which case AM is absent. Regardless of the details of particular algorithms, any window-matching technique basically follows the displacement of the windowed signal, although a single displacement estimate cannot match the displacement at all points throughout the window unless the within-window strain is zero. At some location the displacement estimate precisely matches the actual displacement, provided that displacement estimation error is small. The actual estimation location can be treated as a random variable, with low probability density at window ends and higher probability at the centre. Therefore, a window centre assumption is appropriate if location estimates cannot be refined based on measured signal properties.

Does AM really have any effect, and can it be estimated? Figure 3.1 demonstrates that AM certainly has an effect when displacement estimation is based on phase zero methods. More generally there are good reasons for expecting that no displacement estimator can produce optimal estimates with greatest validity at the window centre, because some signal portions contain little or no information. Figure 3.3 shows simple examples of displacement estimation considering pulse trains. When there is no between-pulse information, an optimal displacement estimator tracks the displacement of the pulse(s) within each window. The example medium in Figure 3.3 has been deformed with a uniform strain, so displacement varies linearly with distance.

An assumption of estimation at the window centre leads to significantly different strain estimates if (a) overlapping windows track the same pulse, or (b) neighbouring windows track pulses at their extremities. When a uniform strain, $\varepsilon$, is tracked in the total absence of displacement estimation error, AM distorts the strain estimates dramatically. The strain estimation lower bound is 0 for overlapping windows, and the upper bound is $\varepsilon \times \frac{Y + \Delta y}{\Delta y}$, where $Y$ is the window length and $\Delta y$ is the window spacing. For non-overlapping windows the lower bound is $\varepsilon \times \frac{\Delta y - Y}{\Delta y}$.

AM could be corrected easily if real ultrasound signals consisted of pulse trains. The correction would entail noting the locations of tracked pulses instead of associating displacement estimates with window centres. Real ultrasound signals are not pulse trains, but they do exhibit large amplitude variations, and portions with lower amplitude usually have lower SNR. Therefore, displacement estimators incorporating considerations of optimality are more responsive to displacements in signal portions with higher amplitude. This implies AM, i.e., the estimation location may be perturbed away from the centre of the window.
3.2.2 AM correction

Location estimation for AMC can be adapted to any displacement estimator based on window-matching. Practical implementation of AMC is in fact effective and very simple, as demonstrated in Section 3.3.

The most difficult task in implementing AMC is finding an approximation to the displacement estimate at window \( n \), \( \hat{u}_n \), expressed as a weighted average of the actual displacements, \( \{u(y)\} \):

\[
\hat{u}_n \simeq \sum_{y=n}^{n+Y} \hat{W}(y, \hat{u}_n) u(y) \left/ \sum_{y=n}^{n+Y} \hat{W}(y, \hat{u}_n) \right.
\]

(3.6)

\( \Delta y \) denotes the window spacing, \( Y \) is the window length, and \( \{\hat{W}(y, \hat{u}_n)\} \) are weighting estimates. \( \{\hat{W}(y, \hat{u}_n)\} \) must be produced in a form that can be evaluated based on properties of the recorded ultrasound signals. Descriptions of practical estimators usually look different to Equation 3.6, because few estimators are implemented directly as weighted averages of point-wise displacement estimates. On first inspection, correlation coefficient, phase zero, SAD and mutual information bear little resemblance to this, but it is possible to obtain suitable approximations by theoretical analysis, or by drawing on empirical experience. For the analytical approach, a signal model may be adopted that neglects noise and decorrelation. Clearly, this analysis is separate from the task of investigating the scale of displacement estimation errors — in the simplified model, tissue deformation results

\[\text{All of this analysis assumes a fairly constant speed of sound, so that length in the axial dimension of ultrasound images corresponds directly to length in physical space.}\]
in an identical warping of the ultrasound signal. Section 3.2.3 demonstrates analytical derivation of a weighting approximation.

This enables location estimation when a linear model of the local displacement field is adopted:

$$u(y) \simeq \alpha + \varepsilon y.$$  \hspace{1cm} (3.7)

Equation (3.7) expresses the assumption that displacement at the scale of each window varies linearly with distance, with strain $\varepsilon$ and offset $\alpha$. This is substituted into Equation 3.6 and rearranged as follows:

$$\hat{u}_n \simeq \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y, \hat{u}_n)(\alpha + \varepsilon y)}{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y, \hat{u}_n)},$$

$$\simeq \alpha + \varepsilon \left[ \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y, \hat{u}_n)y}{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y, \hat{u}_n)} \right].$$ \hspace{1cm} (3.8a)

The location estimate, $\hat{y}_n$, is the location at which the approximation of the displacement estimate equals the actual displacement, i.e., $\hat{u}_n \simeq \alpha + \varepsilon \hat{y}_n$. Hence,

$$\hat{y}_n = \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y, \hat{u}_n)y}{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y, \hat{u}_n)}.$$ \hspace{1cm} (3.9)

In words, location estimates are produced by evaluating the centroid of a set of weighting estimates after windows have been matched for displacement estimation purposes. The main practical undertaking in this chapter amounts to a demonstration and evaluation of results when applying this formula.

The combination of refined location estimates with the same displacement estimates yields greater accuracy overall in the correspondence between the estimates and the actual displacement field. This boosts the accuracy of any technique requiring displacement estimates, not least strain imaging, for example by substituting $\hat{y}_n$ back into Equation 3.1. It also results in more accurate correspondence between tissue features’ physical locations and their apparent locations in strain images.

It should be noted that strain may not actually be uniform within each window, as per Equation 3.7, but the location estimate is based on a linear fit. This is invariably superior to the alternative assumption that within-window strain is zero. Only severe errors in the weighting approximation can render location estimation less accurate than the window centre assumption. The location estimation technique is referred to as AMC. Its application is illustrated in Figure 3.4.

### 3.2.3 AMC for phase zero methods

The derivation of a suitable weighting approximation for phase zero methods is presented. Windows are matched by identifying the zero crossing in the phase of the complex cross-correlation function. Phase zero methods require analytic signals, which are produced...
Figure 3.4: Example of AMC: this signal has been resampled as in Figure 3.1 to simulate a 2% compression. Phase zero estimation is employed to align pre- and post-deformation windows. AMC is applied to refine the location of the displacement estimate. The weightings come from a formula derived in Section 3.2.3. There is perceptible misalignment to either side of the centroid that serves as a location estimate.

by applying the Hilbert transform (or an approximation thereof) to RF ultrasound data. The complex cross-correlation, $\langle a_1, a_2 \rangle$, and correlation phase, $\Phi$, of analytic signals $a_1$ and $a_2$ are

$$\langle a_1, a_2 \rangle (n\Delta y, \tilde{u}) = \sum_{y=n\Delta y}^{n\Delta y+Y} a_1^*(y) a_2(y + \tilde{u}), \quad (3.10a)$$

$$\Phi(n\Delta y, \tilde{u}) = \angle \langle a_1, a_2 \rangle (n\Delta y, \tilde{u}), \quad (3.10b)$$

where $^*$ denotes the complex conjugate, $n\Delta y$ is the beginning of the analysis window in the pre-deformation signal, $Y$ is the window length, and $\tilde{u}$ is a candidate displacement applied to the post-deformation window to look for a match. Eventually the displacement estimate, $\hat{u}_n$, is found at a phase zero.

$$\Phi(n\Delta y, \hat{u}_n) = 0 \quad (3.11)$$

Phase zeros occur once for every wavelength shift in $\tilde{u}$ when $\Phi$ is expressed only in the range $[-\pi, +\pi]$. It is therefore necessary to use tracking strategies such as those described in Chapter 2.

A signal model is adopted that offers a high degree of generality. The pre- and post-deformation analytic signals, $a_1$ and $a_2$, are described as each comprising a common

---

†Later chapters refer back to this model.
component, in addition to independent noise components:

\[ a_1(y) = a_{r1}(y) + ja_{i1}(y), \]  \hspace{1cm} (3.12a)

\[ s_1(y)e^{j\phi_{s1}(y)}, \]

\[ f(y)e^{j\phi(y)} + n_1(y)e^{j\phi_{n1}(y)}, \]  \hspace{1cm} (3.12b)

\[ a_2\left(y + u(y)\right) = a_{r2}\left(y + u(y)\right) + ja_{i2}\left(y + u(y)\right), \]  \hspace{1cm} (3.12c)

\[ s_2\left(y + u(y)\right)e^{j\phi_{s2}(y+u(y))}, \]

\[ f(y)e^{j\phi(y)} + n_2(y)e^{j\phi_{n2}(y)}. \]  \hspace{1cm} (3.12d)

\[ a_{r1} \text{ and } a_{r2} \text{ are the real parts and } a_{i1} \text{ and } a_{i2} \text{ are the imaginary parts of the analytic signals, which can be expressed in phasor notation with envelopes } s_1 \text{ and } s_2 \text{ and phases } \phi_{s1} \text{ and } \phi_{s2}. \]

The model for analysis is expressed in Equations (3.12b) and (3.12d). The pre- and post-deformation signals contain a common component, \( f(y)e^{j\phi(y)} \), warped identically in accordance with the deformation in the underlying tissue, while all other confusion components are grouped in the noise components, with envelopes \( n_1 \) and \( n_2 \) and phases \( \phi_{n1} \) and \( \phi_{n2} \).

In the absence of bulk error, a displacement estimate is similar but not equal to the local displacement at every point in the window. To help with analysing the properties of matched windows, a new symbol is introduced, \( y_2(y, \tilde{u}) \) or \( y_2(y, \tilde{u}_n) \), denoting the pre-deformation location in \( a_1 \) of the datum from \( a_2 \) with which \( a_1(y) \) is matched, \( i.e. \),

\[ y_2(y, \tilde{u}_n) + u\left(y_2(y, \tilde{u}_n)\right) = y + \tilde{u}_n. \]  \hspace{1cm} (3.13)

\( y_2(y, \tilde{u}_n) \) is abbreviated to \( y_2 \) for readability in the remaining equations. The complex cross-correlation function at the match is expressed as follows:

\[ \langle a_1, a_2 \rangle \left(n\Delta y, \tilde{u}_n\right) = \rho_d \left(n\Delta y, \tilde{u}_n\right) + \rho_s \left(n\Delta y, \tilde{u}_n\right), \]  \hspace{1cm} (3.14a)

where

\[ \rho_d \left(n\Delta y, \tilde{u}_n\right) = \sum_{y = n\Delta y} \left\{ f(y)n_2(y)e^{j\phi(y_2) - \phi(y)} + n_1(y)f(y_2)e^{j\phi(y_2) - \phi_{n1}(y)} + n_1(y)n_2(y)e^{j\phi_{n2}(y) - \phi_{n1}(y)} \right\}. \]  \hspace{1cm} (3.14b)

\[ \rho_s \left(n\Delta y, \tilde{u}_n\right) = \sum_{y = n\Delta y} \left\{ f(y)n_2(y)e^{j\phi(y_2) - \phi(y)} \right\}. \]  \hspace{1cm} (3.14c)

This splits the complex cross-correlation function into two parts. \( \rho_d \) contains terms associated with signal stretching, and \( \rho_s \) is associated with noise. Every term in \( \rho_s \) is a sum over the product of signals that are uncorrelated, which tends to cancel out unless \( Y \) is small, while \( \rho_d \) is the informative component. Following the method of Section 3.2.2, \( \rho_s \) is neglected throughout the following analysis to derive a weighting approximation.

The imaginary part of the complex cross-correlation must be zero in order to satisfy the match criterion of Equation (3.11):

\[ \Im \left( \sum_{y = n\Delta y} f(y)f(y_2)e^{j\phi(y_2) - \phi(y)} \right) = 0. \]  \hspace{1cm} (3.15)
This leads to an alternative expression for the phase zero condition:

\[ \sum_{y=n\Delta y}^{n\Delta y+Y} f(y) f(y_2) \sin \left( \phi(y_2) - \phi(y) \right) = 0. \]  

(3.16)

A simple rearrangement of Equation 3.13 gives \( y_2 - y = \hat{u}_n - u(y_2) \), which is the local discrepancy between the displacement estimate and its actual value. In the absence of bulk errors, this is usually small compared to \( \lambda \) throughout any window. It is therefore reasonable to apply the small angle approximation:

\[ \sum_{y=n\Delta y}^{n\Delta y+Y} f(y) f(y_2) \left( \phi(y_2) - \phi(y) \right) \approx 0. \]  

(3.17)

For further simplification, the local mean frequency, \( \bar{\omega}(y, y_2) \), is defined as:

\[ \bar{\omega}(y, y_2) \triangleq \frac{\phi(y_2) - \phi(y)}{y_2 - y}. \]  

(3.18)

The approximation of the zero phase condition can be expressed using \( \bar{\omega}(y, y_2) \):

\[ \sum_{y=n\Delta y}^{n\Delta y+Y} f(y) f(y_2) \bar{\omega}(y, y_2) (y_2 - y) \approx 0, \]  

(3.19)

which can be converted to an expression with clearer relevance to the physical deformation by examining \( y_2 - y \). By expanding a Maclaurin expansion about \( u(y) \), Equation 3.13 can be rewritten as

\[
y_2 - y = \left( \hat{u}_n - u(y) \right) - \left( u(y_2) - u(y) \right),
\]

\[
= \left( \hat{u}_n - u(y) \right) - \frac{du(y)}{dy} (y_2 - y) - \mathcal{O}\left((y_2 - y)^2\right),
\]

\[
= \left( \hat{u}_n - u(y) \right) - \varepsilon \left( \hat{u}_n - u(y_2) \right) - \mathcal{O}\left((\hat{u}_n - u(y_2))^2\right).
\]  

(3.20)

The term scaled by \( \varepsilon \) (strain) and higher order terms are neglected, since all strain imaging is performed with \( \varepsilon \ll 100\% \). The result from Equation 3.20 is substituted into Equation 3.19, giving:

\[ \sum_{y=n\Delta y}^{n\Delta y+Y} f(y) f(y_2) \bar{\omega}(y, y_2) (\hat{u}_n - u(y)) \approx 0. \]  

(3.21)

Rearrangement yields a good approximation for the displacement estimate as a weighted average:

\[ \hat{u}_n \approx \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} f(y) f(y_2) \bar{\omega}(y, y_2) u(y)}{\sum_{y=n\Delta y}^{n\Delta y+Y} f(y) f(y_2) \bar{\omega}(y, y_2)}. \]  

(3.22)

One final modification is made to obtain an approximation that can be applied easily. It is possible to estimate frequency variations within recorded RF ultrasound, but these
usually have low resolution, so it may be difficult to estimate local frequency changes accurately within a single displacement estimation window. However, the nominal fractional bandwidth of medical ultrasound probes is almost always less than 100%, often less than 50%, and frequency variation within individual windows may be smaller still. Therefore, the frequency terms in Equation 3.22 are cancelled, accepting that local frequency variation remains as a source of error.

\[
\hat{u}_n \simeq \frac{\sum_{y=n \Delta y}^{Y} f(y) f(y_2) u(y)}{\sum_{y=n \Delta y}^{Y} f(y_2)}. \tag{3.23}
\]

This is a suitable weighting approximation for implementing AMC with phase zero estimation (cf., Equation 3.6). For practical location estimation, the weightings are evaluated based on the recorded pre- and post-deformation RF ultrasound signals.

\[
\hat{W}(y, \hat{u}_n) = f(y) f(y_2) \tag{3.24a}
\]

\[
= \text{pre-deformation envelope} \times \text{post-deformation envelope} \tag{3.24b}
\]

It might be possible to infer the envelope of the common component, \( f \), separate to the noise component, \( n \). Alternatively, the envelope of the entire measured signal, \( s \), is used as an approximation (cf., Equation 3.12). Location estimation for AMC is then performed by substituting the weightings into Equation 3.9.

### 3.2.4 AMC for correlation coefficient methods

Correlation coefficient methods are widely employed for displacement estimation in ultrasonic strain imaging [13, 46, 81, 145, 164, 166, 167, 168, 176]. \( \rho_{r_1 r_2} \) denotes the correlation coefficient of real RF ultrasound signals \( r_1 \) and \( r_2 \). This is calculated at window \( n \) with candidate shift \( \hat{u} \) following Equation 2.1, with suitable interpolation to estimate correlation coefficients at subsample shifts. A displacement estimate is found at the peak:

\[
\hat{u}_n = \arg \max_{\hat{u}} \rho_{r_1 r_2} (n \Delta y, \hat{u}). \tag{3.25}
\]

The performance of this estimator is similar to phase zero, but not identical. It would be desirable to derive a complementary location estimate following the same analytical procedure. A suitable starting point would be differentiation of \( \rho_{r_1 r_2} \) to examine properties of the maxima, again neglecting the noise components of the pre- and post-deformation signals. This would need to be related to the displacement field to obtain a weighting approximation.

However, an heuristic approach is adopted in this instance, because the derivation would be more complicated than for phase zero, and past experience is indicative of a likely form for the weightings. The significance of each sample as a contributor to the overall correlation coefficient clearly depends on its magnitude, meaning that correct alignment is more significant for high value portions of the signal. For example, if the pulse trains in Figure 3.3 were accompanied by lower amplitude harmonic components
throughout the entire signal, displacement estimates based on correlation coefficient would nevertheless depend almost entirely on the displacements of the pulses. The following heuristic weighting approximation reflects this observation:

\[
\hat{W}(y, \hat{u}_n) = |r_1(y)r_2(y + \hat{u}_n)|. \tag{3.26}
\]

Whether or not these weightings are useful is uncertain given the absence of a theoretical derivation, but it should not be forgotten that the weightings derived in Section 3.2.3 also incorporate approximations, the acceptability of which is not obvious prior to testing.† Simulation results are presented later to investigate the performance of both displacement estimators with AMC based on these weightings.

The type of correlation coefficient considered here, \(\rho_{r_1r_2}\), is the most popular form, but it should be noted that many variants exist. One of the alternatives entails searching for the peak in the real part of the complex correlation coefficient of passband analytic signals, which is more closely related to the phase zero method analysed in Section 3.2.3. Appendix A.2 outlines an analytical argument showing that AMC following Equation 3.24 might be suitable in that case.

### 3.2.5 AM suppression alternatives

The locations of displacement estimates are perturbed away from window centres, particularly by variation in the signal amplitude, but AMC is just one approach to handling this phenomenon.

**Amplitude compression**

If amplitude does not vary within a window, the location of the displacement estimate really may correspond to the window centre. In fact, amplitude variation is not the only signal property that modulates location, but it is certainly the dominant factor. This can be eliminated by pre-processing the signals to ensure uniform amplitude. In phase zero methods this entails scaling the signal to obtain a uniform envelope, while in correlation coefficient methods the amplitude can be compressed to a one-bit signal, recording only whether the value is positive (+1) or negative (−1). According to the above AMC approximations, both of these amplitude compression techniques tie the estimation location to the window centre (cf., Equations 3.9, 3.24 and 3.26).

On the other hand, amplitude compression is not without disadvantages. It entails selectively amplifying signal portions that are weak. If noise and decorrelation are multiplicative, there is no reason to prefer high amplitude portions of the signals, but much of the noise and decorrelation may be additive. This means that amplification of weak signal portions may dramatically reduce the SNR. In turn, this reduces the accuracy of

---

†It may also not be obvious that the phase zero weightings in Equation 3.24 are not applicable to correlation coefficient methods. However, a preliminary comparative test was undertaken, in which the heuristic of Equation 3.26 performed considerably better.
displacement estimation. Amplitude compression nonetheless boosts strain estimation performance where location error is more important than displacement error (cf., Equation 3.5).

Clearly, the value of amplitude compression depends on the relative sizes of displacement and location errors. Sometimes a weaker form of amplitude compression is desirable, to avoid excessive amplification of noise whilst still achieving some suppression of location errors. This can be performed by logarithmic compression of the amplitude, so that fluctuations are reduced, but not eliminated.

Periodic bias errors have been cited as a limitation of amplitude compression [177], because compressed signals are sometimes interpolated with lower accuracy. There is also greater bias in peak interpolation when subsample displacements are estimated from correlation coefficients computed at integer sample displacements. However, these types of error can be eliminated by careful implementation. Peak interpolation is not necessary for subsample estimation. Estimates can be produced instead by signal interpolation to evaluate correlation coefficients directly at subsample displacements. Amplitude compression need not reduce the accuracy of signal interpolation, because it can be applied after rather than before interpolation. This provides a valuable comparison for AMC.

Adaptive stretching

Adaptive strain estimators work on the principle of reversing the deformation that has occurred, usually by stretching the signal in the post-deformation analysis window, so that it maximises a similarity measure with the pre-deformation signal. There is a limit to the number of parameters that can be estimated sensibly, so strain is usually assumed to be constant throughout each window, with an additional search over displacement, thereby determining the two parameters that best align the signals (cf., Equation 3.7). Although its output may still be dominated by the movements of high amplitude signal portions (cf., Figure 3.3), adaptive stretching has the advantage of handling nonuniform displacements throughout each window, in so far as a model assuming uniform strain fits the deformation. Adaptive stretching may not require location estimation, because a stretch estimate may in principle fit displacements correctly throughout the entire window. It is therefore potentially an effective means of suppressing AM.

The performance increase from adaptive stretching is explained in an alternative paradigm as resulting from an increase in the correlation coefficient between the pre- and (stretched) post-deformation signals [178]. It is indeed the case that adaptively stretched windows have higher correlation coefficients. It is not self-evident, however, that increasing the correlation coefficient will always lead to greater accuracy. Adaptive stretching is still subject to error when noise and decorrelation are present.

In any event, the results from past investigations of adaptive stretching have been encouraging [136, 137, 173, 174, 175]. The fact that it is more computationally intensive is usually regarded as its main disadvantage, but adaptive stretching is tested in this chapter as an AM suppression benchmark.
CHAPTER 3. LOCATION ESTIMATION

3.3 Methods

Different strategies for handling AM are compared quantitatively on the basis of SNR\textsubscript{e} measurements (see Equation 3.4). Simulations are employed for assessing the algorithms under controlled conditions. The simulations include noise and strain decorrelation, but exclude decorrelation due to lateral and elevational motion. These phenomena do affect practical scanning, but it is not unreasonable to assume that comparative performance results in the presence of motion decorrelation would be substantially the same, at least in terms of algorithm rankings. Motion decorrelation is essentially just another source of noise, albeit with a different autocorrelation length to white noise.

Algorithm performance is also tested using recorded RF data. Some of the algorithms are applied to data from freehand scanning of \textit{in vitro} and \textit{in vivo} subjects to verify that the expected performance is reproduced under real scan conditions. On the other hand, phantom studies are of limited value for acquiring quantitative results. SNR\textsubscript{e} data are corrupted by nonuniformity in the mechanical properties of the phantom (which is to some degree inevitable) and by nonuniformity in the applied stress field, which is also inevitable except in the hypothetical case of a perfectly uniform axial compression with perfect slip at the upper and lower contact surfaces. Moreover, uniform axial compression causes nonuniform motion in both of the non-axial directions, so the level of decorrelation in any scan is nonuniform. It is therefore difficult in a phantom study to draw definite conclusions regarding the impact of noise and decorrelation on algorithm performance. Notwithstanding these limitations, data are presented from a controlled compression of a phantom with uniform mechanical and scattering properties.

Many filtering techniques exist for suppressing the translation from displacement noise through to strain estimation noise, but no post-filtering is applied in the following tests, so as to obtain a clear picture of displacement estimation performance. Strain is estimated by differencing displacement estimates that are closely spaced, so changes in the noise level are clearly visible and measurable. The remainder of this section comprises descriptions of algorithm implementations, followed by details of the simulations, the phantom study and freehand scanning.

3.3.1 Algorithms

The tests cover variations on phase zero, correlation coefficient and adaptive strain estimation. Phase zero and correlation coefficient are tested both in basic implementations and with combinations of amplitude log compression, limiting log compression and AMC. In order to compare performance fairly, the window length is fixed across all of the estimators in each test. The window spacing is fixed at $\Delta y = 2.7\lambda$ (\textit{i.e.}, 0.45 $\mu$s, 0.35 mm, 30 samples), with strain calculated using Equation 3.1, so changes in displacement and location error can be measured easily (\textit{cf.}, Equation 3.5). For consistency, the same signal interpolation is employed with every algorithm.
Efficient phase zero search

The efficient phase zero search (EPZS) is adapted from [80], with a simplification noted in the preliminary study [14], and increased robustness due to the use of cross-seeding from Chapter 2. Techniques of this sort represent iterative extensions of older, usually non-iterative phase methods that first emerged in medical ultrasound for pulsed Doppler velocity estimation [179].

To summarise, matched 5–10 MHz filters are applied to the RF signals to produce analytic signals, \(a_1\) and \(a_2\), which are converted to baseband signals, \(a_{b1}\) and \(a_{b2}\), using a modulation frequency, \(\omega_0\), which is chosen approximately equal to the centre frequency. Subsample values in \(a_{b2}\) are obtained by baseband linear interpolation, enabling accurate subsample displacement estimation \(\hat{u}_{n}\). Phase zero methods require that the displacement of each window is known already to within \(\lambda/2\), hence the need for robust tracking to provide the initial seed, \(\tilde{u}_0\). The following, shortened iteration formula arises due to the baseband conversion [14]:

\[
\tilde{u}_{k+1} = \arg\langle a_{b2}, a_{b1} \rangle (n\Delta y, \tilde{u}_k) \frac{\omega_0}{\Delta y}.
\]

(3.27)

Here the arg values are wrapped to the range \([\tilde{u}_k\omega_0 - \pi, \tilde{u}_k\omega_0 + \pi]\). The search converges, terminating when \(\tilde{u}_{k+1}\) and \(\tilde{u}_k\) differ by less than a threshold (e.g., 0.001 samples).

The location estimate, \(\hat{y}_n\), for strain estimation in Equation (3.1) is usually assumed to be the window centre:

\[
\hat{y}_n = n\Delta y + \frac{Y}{2}.
\]

(3.28)

In tests with amplitude log compression, a new baseband signal, \(a_{b,\log}\) is given by:

\[
a_{b,\log}(y) = \log\left(1 + c|a_b(y)|\right)e^{j\arg a_b(y)}.
\]

(3.29)

c is the compression factor. The larger the value of c, the smaller the amount of amplitude information that is retained. This algorithm is referred to as EPZS_L1.

Eventually, as \(c \to \infty\), all of the amplitude information is discarded. This is referred to as limiting log compression, EPZS_L2, which is the phase zero counterpart of one-bit compression in correlation coefficient methods [168, 169, 181]. It has the form:

\[
a_{b,\log}(y) = e^{j\arg a_b(y)}.
\]

(3.30)

The other tests use AMC. EPZS combined with AMC is denoted by EPZS_A. When producing baseband signals, the envelope is also detected, \(s_1(y)\) (cf., Equation 3.12), which is used to obtain AMC weightings for location estimation in Equation 3.9:

\[
\hat{W}(y, \hat{u}_n) = s_1(y)s_2(y + \hat{u}_n).
\]

(3.31)

Additionally, since amplitude information is not entirely absent in EPZS_L1, an estimator combining this with AMC is denoted by EPZS_LA, in which weightings are calculated using the log compressed envelope.

\(^1\)Cross-seeding requires a correlation coefficient for every displacement estimate. The complex cross-correlation function is already calculated in EPZS. The real part divided by \(\sum_{y=n\Delta y}^{n\Delta y+Y} s_1(y)^2 + s_2(y + \hat{u}_n)^2\) provides a correlation coefficient.

\(^2\)For discussion of interpolation frequency responses, see [180].
Correlation coefficient maximiser

The correlation coefficient maximiser (CCM) is implemented here for accuracy rather than speed. Analytic signals are calculated as for EPZS. To determine a displacement at each window, the CCM initially searches at integer sample displacements as described in Chapter 2 (with cross-seeding) to find the maximum value of the correlation coefficient, calculated per Equation 2.1. The estimate is refined by allowing subsample values of $\tilde{u}$. At this stage, correlation coefficients are calculated with subsample values in $r_2$ found by interpolation. A complex baseband representation of $r_2 (a_{\omega_2})$ enables highly accurate subsample interpolation, although in CCM the baseband result must be translated back to a real passband value. This is performed by the following calculation, where $\omega_0$ is again the modulation frequency used for translating between passband and baseband:

$$r_2(y) = \Re\{a_{\omega_2}(y)e^{j\omega_0 y}\}.$$  \hspace{1cm} (3.32)

$\hat{y}_n$ is still usually assumed to be the window centre, as per Equation 3.28.

Log compression is denoted by CCM_L1, using the formula proposed by Céspedes and Ophir [169]:

$$r_{\log}(y) = \log(1 + c|r(y)|)\text{sign}(r(y))$$  \hspace{1cm} (3.33)

To maximise algorithm performance, baseband linear interpolation is used as before, and $r_2$ is only log compressed afterwards at the point of computing the correlation coefficient. The limiting case for log compression of the real signal is one-bit compression:

$$r_{\log}(y) = \begin{cases} +1 & r(y) \geq 0, \\ -1 & r(y) < 0. \end{cases}$$  \hspace{1cm} (3.34)

Subsample interpolation is still performed by linear interpolation of the uncompressed baseband signal, so zero crossings are identified with high accuracy. This variant is denoted by CCM_L2.

The CCM variant with AMC is CCM_A, which tests the weighting approximation of Equation 3.26. AMC is applied to non-limiting log compression (CCM_L1) in the final variant, CCM_LA.

Adaptive strain estimator

Adaptive strain estimators (ASE) with a range of similarity measures and different stretching methods were tested to find the right benchmark. Appendix B provides descriptions, discussion and quantitative comparisons. The particular ASE in the main text is based on a semi-exhaustive search over displacement and stretch, using correlation coefficient as the similarity measure.
3.3.2 Simulation

Simulations offer precise control over echogenicity (scatterer strength and scatterer density) and deformation (transformation of scatterer positions between consecutive frames). Changed spacing of the scatterers causes decorrelation, as in real ultrasound scanning, and random noise can be added to make simulations more realistic. The addition of noise is particularly important for simulations with strains below 0.5%, where strain decorrelation is relatively low, so background electrical noise is relatively significant.

The type of scattering field is obviously an important consideration. The echogenicity could be made to vary over small distances to exaggerate AM and introduce artefacts in algorithms without AMC or AM suppression. On the other hand, the existence of AM artefacts has already been demonstrated, whereas it is more interesting to examine the effect of AM on strain estimation noise in homogeneous scattering fields, where its effect is not obvious.

RF ultrasound data from uniform scattering fields were simulated using Field II \cite{147}. The simulations have $2 \times 10^5$ scatterers distributed uniformly throughout $50 \times 50 \times 6$ mm volumes with scattering strengths also uniformly distributed. Axial strains were simulated by rescaling the axial coordinates of the scatterer positions between consecutive frames.

The simulated probe parameters model the 5-10 MHz probe of the system described in Section 1.4, since this is used for freehand scanning in other experiments. The sampling frequency is 66.7 MHz, and experimental measurements with the probe indicated a centre frequency of 6.0 MHz with a 2.1 MHz bandwidth.

Data were simulated for five independent scatterer fields with the above statistical properties, each having 128 A-lines spanning 40 mm in the lateral direction, recorded to a depth also of 40 mm. The compressive strains were 0%, 0.01%, 0.1% 0.5% 1.0%, 1.5%, 2.0%, 3.0% and 4.0%, offering a good indication of the strain dependency of algorithm performance.

Simulation data were converted to the format of the Stradwin\footnote{http://mi.eng.cam.ac.uk/~rwp/stradwin} freehand 3D ultrasound system before strain estimation. Samples were recorded with 16-bit signed integer precision. The data were normalised to give a root mean squared signal strength of $2^{10}$, corresponding to a mean SNR of 71 dB in the presence of quantisation noise. Most tests were performed with the addition of white noise, reducing the SNR to 20 dB. The top and bottom 3.5 mm in each image is omitted when evaluating SNR, because the range of window lengths is restricted at the edges. Therefore, every numerical result is an average over strain estimation in five independent simulation frames, 128 A-lines per frame and 95 strain estimates per A-line.

3.3.3 Phantom

An agar phantom was constructed with uniform mechanical and scattering properties (85.7 wt.% water, 10.9 wt.% glycerol, 1.7 wt.% agar M1002\footnote{http://www.melford.co.uk} 0.85 wt.% Biocide\footnote{http://www.fisher.co.uk} and
0.85 wt.% aluminium oxide powder) forming a cylinder of height 40 mm and radius 30 mm. Scanning was undertaken using the system described in Section 1.4. The probe was mounted in a mechanical rig for precise control of vertical (i.e., axial) movement above the phantom, and a footprint extender was attached to facilitate uniform compression over the entire top surface. Ultrasound frames were acquired before and after a 1.2% compression. In the top and bottom bands of the images the data are relatively nonuniform owing to reverberation at the face of the ultrasound probe and imperfect slip at the phantom boundaries, but strain estimation was performed using many of the algorithms in this paper, recording SNR values for the middle band at 10–30 mm depth, where the results are more reliable.

### 3.3.4 In vitro and in vivo scanning

Freehand scans were performed using the same equipment as the phantom study, but the probe was taken out of the mechanical rig and the footprint extender was removed. As per previous work [14], frames were acquired at $\sim30$ Hz during each scan, and exaggerated palpating movements were not necessary. The images are used to verify that the performance indicated by simulations is reproduced when analysing real scan data.

### 3.4 Results

The quantitative results from analysis of the homogeneous simulations are presented first. These indicate the effect of AM on general estimation error, without distorting the comparison of different algorithms by analysing extreme cases. A systematic presentation of results across ranges of parameter settings and strains illuminates exactly what AM entails and how different algorithms handle it. The phantom study produces similar results, verifying that the simulations are a good model of real ultrasound data.

It would be wrong, on the other hand, to assume that real strain images are not also affected by artefacts due to inhomogeneous echogenicity. When severe AM artefacts occur it can become impossible to interpret strain images correctly. In vitro and in vivo images are therefore valuable for an appreciation of the significance of AM for practical scanning.

#### 3.4.1 Window length

The first set of results shows how changes in window length impact differently on algorithm performance depending on whether or not AMC is applied. The graphs in Figures 3.5 and 3.6 show at a moderate strain how AMC changes both EPZS and CCM, with the ASE benchmark plotted for comparison. Window lengths span the range $2.8–27.1\lambda$.

Further results in Figure 3.7 focus on the performance of ASE at higher strains, where it is most advantageous. Although algorithms with AMC perform similarly or better when using short windows, without stretching they exhibit a drop in performance when using long windows, because they are eventually subject to within-window phase-wrapping. On
Figure 3.5: \( \text{SNR}_e \) against window length for EPZS and EPZS_A, with both 71 dB and 20 dB data at 0.5% strain. Uncorrected EPZS with 71 dB data reaches a plateau at \( Y = 10\lambda \), which the 20 dB results converge towards for long windows. When AMC or ASE is applied there is no such plateau and much higher \( \text{SNR}_e \) is achieved — \( \text{SNR}_e \) is almost a linear function of window length, indicating that location error in the corrected algorithms is negligible.

On the other hand, it is often desirable to avoid long windows anyway so as to achieve high resolution.

Figure 3.8 shows example strain images to demonstrate the implications for image quality of changes in \( \text{SNR}_e \). The images can be compared with the corresponding \( \text{SNR}_e \) results from the graphs. Brightness in the images is proportional to strain: black is zero strain, mid-grey is the simulated strain and white is twice the simulated strain, with saturation at both ends of the scale. The images were constructed by nearest-neighbour interpolation without any filtering. An ideal estimator would yield a uniform greyscale level, but noise introduces inhomogeneity.

### 3.4.2 Compression factor

A fixed window length was employed to test various levels of amplitude log compression. The use of very long windows imposes a large penalty on algorithms that incorporate neither AMC nor AM suppression. For these reasons, \( 13.5\lambda \) is employed for the remaining quantitative results.

Performance with log compression is recorded against different values of the compression factor that determines compression strength (cf., Equation 3.29 and Equation 3.33). The effect of log compression varies widely depending on the strain, so Figures 3.9–3.11 show results at the smallest, mid-range and largest strains from the simulations. In interpreting these graphs, it should be noted that \( \text{SNR}_e \) is proportional to the strain, so low
values of SNR\(_e\) in Figure 3.9 do not imply that the level of estimation noise was higher than in the other tests (it was actually lower). These data are most interesting as a verification of the theory in Section 3.2 regarding the two sources of estimation noise. Nevertheless, the quantitative data show that log compression is not desirable at all strains.

### 3.4.3 Strain dependence

The final simulation tests mostly continue with a window length of 13.5\(\lambda\), and the compression factor is fixed at \(c = 10^3\). This is chosen at a relatively high value to give good performance at high strains, since it is at high strains that log compression has the largest positive effect compared to algorithms without AM suppression.

The final simulation graphs show the performance of all of the algorithms across the full range of simulated strains. Figure 3.12 compares all members of the EPZS family, Figure 3.13 covers the CCM family, and Figure 3.14 is a comparison between ASE and the displacement estimators with AMC. Figure 3.14 also compares SNR\(_e\)-strain characteristics at a shorter window length, highlighting the effect of phase-wrapping.

### 3.4.4 Phantom study

The phantom data at 1.2% compression exhibit very similar trends to the simulations. SNR\(_e\) is plotted in Figure 3.15 against window length for a large set of algorithms. Comparison with Figures 3.5–3.6 shows that the simulations provide an accurate guide to the
Figure 3.7: SNR$_e$ against window length for EPZS$_A$, CCM$_A$ and ASE, with 20 dB data at 4% strain. ASE performs less well with short windows, but the algorithms without stretching peak at roughly 10$\lambda$ and performance decreases after that point, whereas ASE continues to perform better with longer windows. The drop in performance at window lengths $>$10$\lambda$ is caused by the large displacement difference $>$0.4$\lambda$ between the ends of the windows, which leads to phase-wrapping errors, as discussed in Section 3.5.

Figure 3.8: Examples from simulations: These strain images were generated from 20 dB data at 0.5% compression using $Y = 13.5\lambda$ with EPZS, CCM, EPZS$_A$ and CCM$_A$. The performance of EPZS and CCM is similar, though EPZS$_A$ outperforms CCM$_A$.

Although the performance of ASE is relatively erratic, it seems unlikely that this indicates any fundamental problem with adaptive stretching. Rather it may indicate that, despite measures detailed in Appendix B, it is challenging to implement the necessary multi-parameter search sufficiently robustly to achieve a stable value of SNR$_e$ when processing real scan data (see [175] for further discussion).
Figure 3.9: $\text{SNR}_e$ results for EPZS_L1, EPZS_LA, CCM_L1 and CCM_LA with 20 dB data at 0.01% strain as a function of $c$, the compression factor. At low strains, the main effect of log compression is increased displacement estimation noise. The effect is more pronounced with the CCM family. The addition of AMC had no effect in this test.

Figure 3.10: $\text{SNR}_e$ results for EPZS_L1, EPZS_LA, CCM_L1 and CCM_LA with 20 dB data at 0.5% strain as a function of $c$, the compression factor. At this strain, log compression significantly improves the performance of EPZS. CCM is also improved by slight log compression. Better performance is produced by adding AMC, although as $c \to \infty$ EPZS_LA and CCM_LA converge with EPZS_L1 and CCM_L1, and perform worse than EPZS_A and CCM_A (see Figures 3.5 and 3.6).
Figure 3.11: SNR<sub>c</sub> results for EPZS<sub>L1</sub>, EPZS<sub>LA</sub>, CCM<sub>L1</sub> and CCM<sub>LA</sub> with 20 dB data at 4% strain as a function of c, the compression factor. At this strain all of the algorithms can be improved by applying an appropriate level of log compression. The greatest improvement is exhibited by EPZS<sub>L1</sub>, while the AMC algorithms are still degraded by high compression factors, resulting in performance convergence as in Figure 3.10.

Figure 3.12: SNR<sub>c</sub>-strain characteristics for the EPZS family of algorithms with 20 dB data. EPZS<sub>A</sub> has the best performance across a wide range of strains, although the SNR<sub>c</sub> is lower at high strains and at 4% the best performance is from EPZS<sub>LA</sub>. 
Figure 3.13: SNR<sub>e</sub>-strain characteristics for the CCM family of algorithms with 20 dB data. At all strains CCM<sub>A</sub> significantly outperforms the other algorithms. In the absence of AMC, log compression boosts CCM performance at strains above 1.5%, where the best log compression algorithm is CCM<sub>LA</sub>.

Figure 3.14: SNR<sub>e</sub>-strain characteristics for EPZS<sub>A</sub>, CCM<sub>A</sub> and ASE with 20 dB data. These are the best algorithms from each of the three families. EPZS<sub>A</sub> performs best at low and moderate strains, but ASE is the best performer when long windows are applied to data recorded at a high strain. The window lengths are (a) 13.5 λ and (b) 6.0 λ.
3.4.5 In vitro and in vivo results

The final group of results are strain images where selected algorithms are applied to data from real scans of in vitro and in vivo subjects. The brightness scale has black corresponding to zero strain and white corresponding to twice the image mean strain, which is stated in each caption. A small set of algorithms is more informative than a presentation covering all 11 algorithms that were tested. A single ASE image is presented, to remind the reader that ASE exists as an option for producing high quality strain images at extremely high computational cost, but Figures 3.16–3.19 focus on images from three algorithms in the EPZS family.

The first algorithm is plain EPZS, showing what AM entails for performance where it is neither corrected nor suppressed. Secondly, EPZS_L2 represents the behaviour of amplitude compression without ambiguity regarding the choice of $c$ — non-limiting log compression achieves a compromise between EPZS and EPZS_L2, so the reader may choose to imagine intermediate images between those that are included in the figures. Similarly, EPZS_A is chosen as the third algorithm to demonstrate AMC. The compromise of applying AMC together with non-limiting log compression is also a possibility (i.e., EPZS_LA). Again, readers may imagine intermediate images between EPZS_L2 and EPZS_A. These images show that AM occurs in real data, and that different types of correction and suppression produce images with different qualities.

The relative behaviour of the algorithms depends on strain and window length, with the most striking differences when using long windows. Comparative images at $Y = 4\lambda$ illustrate the behaviour with short windows. Other images were produced using $Y = 30\lambda$, which better highlights differences in behaviour. Strain was estimated by differencing windows at a spacing of $2.7\lambda$, as in the quantitative tests. There was still no post-filtering, but the images were produced by linear interpolation and strain was estimated at intervals of $1.35\lambda$ along each A-line for a smoother display.

Furthermore, when AMC is applied to inhomogeneous data, the spacing of estimation locations is found to be uneven. In images with little or no further filtering, this introduces textural variation, which complicates the differences in appearance between images with and without AMC. The textural variation has been avoided in the examples of EPZS_A images by using AMC to place the windows irregularly, so as to achieve regular spacing of the estimation locations. This is done at fairly minimal computational cost by accumulating the location error as samples are added to the window around the required location, adding each next sample to whichever end of the window opposes the current location error.
Figure 3.15: Agar phantom, 1.2% strain, SNR against window length: Measurements for numerous algorithms indicate that the trends observed in the simulations are reproduced when analysing real scan data. Note that although the results for uncorrected CCM appear to be missing, this is because they overlap almost precisely with uncorrected EPZS. Strain images on the right are at the maximum window length (27.1λ).

Figure 3.16: Olive/gelatin phantom (0.28% relaxation): (a) EPZS, (b) EPZS_L2, (c) EPZS_A, with 1: \( Y = 4\lambda \) and 2: \( Y = 30\lambda \).
CHAPTER 3. LOCATION ESTIMATION

Figure 3.17: Inhomogeneous gelatin/gelatin phantom (0.61% compression): (a) EPZS, (b) EPZS_L2, (c) EPZS_A.

Figure 3.18: Transverse scan of forearm (0.52% relaxation): (a) EPZS, (b) EPZS_L2, (c) EPZS_A.

Figure 3.19: Longitudinal scan of thigh (0.68% relaxation): (a) EPZS, (b) EPZS_L2, (c) EPZS_A.
3.5 Discussion

The striking implication of the quantitative results is that AM handling is an important consideration for algorithm design to minimise estimation error in general, even when the scatterer fields are substantially homogeneous. Location rather than displacement estimation noise is the limiting factor on the performance in Figures 3.5 and 3.6. This can be inferred because displacement estimation noise depends on the ultrasonic SNR, but as the window length became large the performance of EPZS and CCM was the same at 20 dB and 71 dB. Performance of the corrected phase zero algorithm, EPZS,A, was far higher, and it depended on ultrasonic SNR regardless of the window length. It is evident that the form of AMC in EPZS,A is slightly more accurate than the heuristic weighting approximation in CCM,A. However, CCM,A also performs substantially better than the uncorrected algorithm, and still outperforms ASE for short and medium window lengths. This indicates that there is considerable scope for heuristic corrections, even though some algorithms may not lend themselves neatly to the analytical method for deriving AMC weightings. On the other hand, if it is not possible even to devise a suitable heuristic, the results show that adaptive stretching produces performance that is almost as good at low strains, and exceeds the performance of AMC when applied with long windows at high strains. AM theory can largely be ignored when implementing an algorithm such as ASE.

Of course, adaptive stretching as in ASE is not the only form of signal warping that can be applied in strain imaging systems. One alternative is global stretching, in which the image mean strain is first estimated, and the entire post-deformation ultrasound frame is stretched uniformly to compensate. A displacement search is then carried out as normal, usually based on the correlation coefficient. By decoupling the fine scale estimation from the signal warping, this may sometimes outperform adaptive stretching. However, global stretching is obviously subject to AM in every region that deviates from the mean strain, as in any scan of inhomogeneous tissue. Therefore, the combination of AMC or log compression with a coarsely varying stretch could be an interesting extension for fast, high quality estimation of high strain deformation data.

The quantitative results with log compression are useful primarily because they support the AM theory. The graphical results demonstrate exactly the behaviour expected from algorithms in which increasing the compression factor brings an increase in displacement error, but reduces location error. The optimal compression factor depends on the relative sizes of the two types of error.

All of the log compression results including the SNR e-strain characteristics indicate that AMC is a better method for handling AM, at least where there are few bulk errors. For AMC with perfect location estimation this is expected, since AMC already eliminates location error without increasing the displacement error. By contrast, assuming that noise has an additive component, log compression reduces location error at the expense of increased displacement error. The exception to this arises at high strains, or when extremely long windows are used. It was shown in Figure 3.7 that EPZS,A and CCM,A eventually have performance reductions when the windows are long. This results from
phase-wrapping, and it affects any estimator that does not stretch the signal. Samples at the end of the window contribute noise if they are misaligned by more than $\lambda/2$. In algorithms that estimate displacement at the centre, such as log compression, phase-wrapping only happens when the total displacement over the length of the window exceeds $\lambda$. However, it sometimes happens earlier in algorithms that estimate away from the window centre, such as EPZS, EPZS_A, CCM and CCM_A (cf., Figure 3.4).

The quantitative results show that AM can be handled quite effectively in EPZS simply by applying log compression, but this does not apply in the case of CCM, as previously noted in [169]. The advantage of applying log compression to EPZS is that compression of the envelope retains the phase, so less information is discarded. Therefore, AMC may in fact be of greater significance when using CCM, even though the absolute performance of CCM_A is lower than EPZS_A, because users of correlation coefficient techniques operating on real RF data have not previously had a fast method for achieving accurate strain estimation free of AM.

In general, it may be surprising to some readers that AMC represents a real-time technique producing somewhat more accurate displacement estimates than ASE in most circumstances. On the other hand, the limitation exposed in the SNR_strain characteristics of Figures 3.12-3.14 is that, in the absence of any stretching, the chosen window length must be sufficiently small to avoid large within-window displacement differences causing phase-wrapping. While this clearly constrains parameter selection for high strain work, results such as Figure 3.15 show that, for low and moderate strains, the window length is effectively unconstrained. This is all the more encouraging considering that many applications (e.g., typical freehand strain imaging of soft tissues at a reasonable frame rate) operate with inter-frame strains below 1%. The reason for the high performance of AMC is that accurate weighting approximations lead to highly accurate location estimates, whereas ASE is independent of AM only if the signal is stretched by a factor corresponding precisely to the true strain. ASE is therefore less accurate when the strain is too low or the windows are too short, in which cases significant errors in the stretch factors mean the algorithm fails to eliminate AM.

Finally, the images in Figures 3.16-3.19 show clearly that AM theory applies to real scan data, so AMC is a valuable technique. The upper set of images in Figure 3.16 shows that AM is relatively unimportant when using short windows — indeed the only perceptible difference between the images is that EPZS_L2 is marginally noisier. The difficulty with interpreting the other images for scientific purposes is that, unlike in simulations, there is no ground truth. However, certain general observations can be made. Firstly, it is clear that long windows cannot be used in algorithms that do not handle AM. All of the objects in EPZS and CCM images end up badly misregistered. Additionally, EPZS and CCM are generally noisier, even in homogeneous regions, than algorithms employing either log compression or AMC. Regarding the means of noise reduction in EPZS_L2 and EPZS_A, the main difference is that, whereas EPZS_A is sensitive to amplitude differences, and preferentially uses data from brighter regions, increasing the window length with EPZS_L2 has a more uniform smoothing effect, so EPZS_A usually achieves lower es-
CHAPTER 3. LOCATION ESTIMATION

...tion error. Where different echogenicities correspond to different stiffnesses, EPZS_A
also substantially avoids blurring the boundaries that are lost in EPZS_L2, as with the
lower regions in Figures 3.16 and 3.17. On the other hand, the advantage of more gen-
eral blurring is that EPZS_L2 is less susceptible to occasional bulk error when bright
regions decorrelate, of which there are some examples in Figure 3.19. This could be prob-
lematic if it occurs often in a particular application, which motivates work on nonlinear
post-processing for outlier detection and removal [139].

3.6 Conclusion

This chapter began with a description of the AM effect with theory to characterise it,
leading naturally to a general correction method: AMC. The correction has been applied
to phase zero and correlation coefficient algorithms. Alternative AM suppression methods
(amplitude log compression and adaptive stretching) were also discussed. Simulation
results from homogeneous scattering fields showed that AMC substantially improves the
accuracy of displacement estimation for strain imaging.

One of the most important points is that AMC entails just one calculation after each
displacement is estimated, so it is suitable for real-time processing. There are choices to
be made regarding exactly how the refined location estimates are applied, but the addition
of location estimation to the processing cost of strain imaging systems can probably be
neglected for the vast majority of applications.

AMC can be regarded as one step among many towards improved displacement esti-
mation by window-matching techniques. In some circumstances, AMC in isolation boosts
displacement estimation performance substantially. In other situations, especially with
very high strains and long windows, some combination of AMC, log compression and
simple warping schemes such as global stretching can potentially achieve lower estimation
error at acceptable computational cost.

The theory and experiments in this chapter refer exclusively to 1D displacement es-
timation with 1D (single A-line) windows. Multi-dimensional theory has been omitted
for concision, but the principle of Section 3.2.2 can be extended trivially to 2D or 3D
windows to improve 1D, 2D or 3D displacement and strain estimation.
Chapter 4

Phase-based estimators

4.1 Background

Chapter 3 showed that the fine-scale accuracy of window-matching approaches to deformation estimation can be improved at little computational cost by AMC: a single calculation is performed to estimate the location at which each displacement estimate applies with greatest validity. This is significant when using windows of finite size, in which there is substantial variation between the displacements at different points in the window. AMC is particularly accurate when displacement estimation is performed by identifying the displacement at which the phase of the complex cross-correlation function crosses zero, as in efficient phase zero search (EPZS).

In the light of this result, further analysis now focuses on phase-based estimators. A new and potentially highly versatile family of deformation estimators is introduced, called weighted phase separation (WPS), while also elaborating on the properties of EPZS. Both algorithm families are compared in an experimental framework similar to that employed in Chapter 3. This is aimed at developing a useful functional description of how phase-based deformation estimators work, identifying strengths and limitations. Additionally, WPS provides a general framework for phase-based deformation estimation, which may offer advantages with regard to accuracy, configurability, and optimisation.

4.2 Theory

WPS is motivated by interest in devising a general approach to displacement estimation, which is particularly suitable for use with AMC, and offers a range of options for tuning the algorithm as knowledge is accumulated regarding the properties of ultrasound signals from deformed scatterer fields. Obvious novel aspects of WPS are not calculating any form of cross-correlation, and operating on data in an envelope/phase (or just phase) representation, rather than real or analytic representations of the RF ultrasound signal. Throughout this thesis, including this chapter, lateral motion is only allowed for by the coarse lateral tracking described in Chapter 2 to select appropriate A-lines. WPS is
presented as a method for fine-scale axial deformation estimation, although the scope for extending the method should become evident.

4.2.1 Weighted phase separation

Relationships noted during the derivation of a weighting approximation in Chapter 3 suggest that signal phase separation could be considered as a stand-alone displacement estimator (see Equations 3.18 and 3.20), without recourse to cross-correlation. If phase differences are expressed in the range \([-\pi, +\pi]\), pre- and post-deformation data must be aligned to within \(\lambda/2\) by a tracking method (see Chapter 2) so as to avoid phase-wrapping errors. Rearrangement of Equations 3.18 and 3.20 gives an approximation for the displacement at position \(y\),

\[
\frac{\hat{u}(y) \simeq \hat{u}_n + \frac{\phi(y) - \phi(y_2)}{\hat{\omega}(y, y_2)}},
\]

where the symbols retain their previous meanings: \(\hat{u}_n\) is a displacement estimate used for aligning the data, and \(\hat{\omega}(y, y_2)\) is the rate of change in phase of the common component of the ultrasound signals (cf., Equation 3.12) at the point of interest.

This can be reinterpreted as a formula for producing a point-wise displacement estimate, \(\hat{u}(y, \tilde{u}_k)\), based on a trial alignment, \(\tilde{u}_k\):

\[
\hat{u}(y, \tilde{u}_k) = \tilde{u}_k + \frac{\hat{\phi}(y) - \hat{\phi}(y_2)}{\hat{\omega}(y, y_2)},
\]

where now \(\hat{\phi}(y)\) and \(\hat{\phi}(y_2)\) are estimates of the common component phase within the pre- and post-deformation data, and \(\hat{\omega}(y, y_2)\) is an estimate of \(\hat{\omega}(y, y_2)\). Once again, a constant value for \(\hat{\omega}(y, y_2)\) is assumed (replacing it with the nominal probe centre frequency, \(\omega_0\)) although this might be worth reconsidering in systems with high fractional bandwidth. Deviation from the centre frequency biases point-wise estimates, although the bias is eliminated later in this analysis.

To estimate the phase separation, \(\hat{\phi}(y) - \hat{\phi}(y_2)\), there may be scope for sophisticated adaptive filtering approaches to suppress the noise component, but in this investigation it is replaced with the phase of the overall signal.

\[
\hat{\phi}(y) = \arg a_1(y) = \phi_{s1}(y) \\
\hat{\phi}(y_2) = \arg a_2(y + \tilde{u}_k) = \phi_{s2}(y + \tilde{u}_k)
\]

Recall that the overall phase of an analytic signal is evaluated in the range \([-\pi, +\pi]\) by taking the inverse tangent of the ratio between its real and imaginary parts, while noting the correct quadrant based on their signs.

\[
\phi_s(y) = \arg a(y) = \tan^{-1}\left(\frac{a_i(y)}{a_r(y)}\right)
\]
Assuming alignment correct to within $\lambda/2$, one might envisage detecting the phase of the RF ultrasound signals, and immediately applying Equation \ref{eq:4.2} to produce point-wise displacement estimates at every sample. However, this suffers from extremely high noise, because there is no chance for noise terms to cancel out. Additionally, the larger the misalignment of the individual points, the lower the accuracy of the approximation in Equation \ref{eq:3.20}, and consequently in Equation \ref{eq:4.2}. Thirdly, error arising due to perturbation of the local frequency away from the nominal frequency is proportional to the misalignment, which makes displacement estimates biased.

Each point-wise displacement estimate can be refined by iterative realignment. The alignment could get worse if this were performed simply using the current point-wise displacement estimate to set the realignment of that point for the next iteration. A more robust approach is to evaluate a weighted average of point-wise estimates across a wider region (\emph{i.e.}, a window) to realign the entire window:

$$
\hat{u}(y, \tilde{u}_k) = \tilde{u}_k + \frac{\phi_{s1}(y) - \phi_{s2}(y + \tilde{u}_k)}{\omega_0},
$$

\begin{equation}
\hat{u}_{k+1} = \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y) \hat{u}(y, \tilde{u}_k)}{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y)}, \tag{4.5b}
\end{equation}

$$
\hat{u}_{k+1} = \hat{u}_k + \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y) \left( \phi_{s1}(y) - \phi_{s2}(y + \tilde{u}_k) \right)}{\omega_0 \sum_{y=n\Delta y}^{n\Delta y+Y} W(y)}. \tag{4.5c}
$$

Point-wise estimates at alignment $\hat{u}_k$ are calculated according to Equation \ref{eq:4.5a}, and Equation \ref{eq:4.5c} is a weighted average for iterative realignment. The point-wise estimates become more accurate with each improvement in alignment, with iterations ceasing like EPZS upon satisfying a convergence criterion (\emph{e.g.}, $\tilde{u}_{k+1}$ and $\tilde{u}_k$ differ by less than 0.001 samples). Further stages of finer realignment could then be pursued, analogous to adaptive stretching. Alternatively, the final window alignment can be recorded as a robust displacement estimate, $\hat{u}_n$. This algorithm is referred to as WPS.

Deviation from the centre frequency is no longer a source of bias in the final displacement estimate, because the numerator on the right-hand side of Equation \ref{eq:4.5c} is zero after convergence. This feature is shared by EPZS, which converges on an alignment at which the phase of the complex cross-correlation function is zero \cite{14,80}. However, WPS has some other interesting properties: (1) WPS is an ideal target for AMC since the displacement estimate is explicitly evaluated as a weighted average, so the same weightings are used for location estimation. For other algorithms, such as EPZS, the weightings need to be estimated. By contrast, with WPS the strategy for assigning weightings is an important aspect of algorithm design. (2) An arbitrary form of within-window phase-unwrapping could be applied in WPS, \emph{e.g.}, to avoid discontinuity in the trend of point-wise displacement estimates in the window. For correlation phase, the effective phase-wrapping is fixed. The version of WPS that is tested here uses a fixed wrapping range for each window, but later results indicate possible benefits if within-window phase-unwrapping is considered in the future. (3) WPS and EPZS entail different calculations operating on
different representations of RF ultrasound data, so they present different opportunities for software optimisation and hardware implementations. Software optimisation described in Appendix C produces similar efficiency for both algorithms, but that need not generally be the case.

### 4.2.2 Weighting selection

Weightings, \( W(y) \), should be chosen to emphasise signal portions that are of special interest. For example, a weighting of zero is implicitly applied to data outside each window. Within windows, it is simplest to apply uniform weightings, \( W(y) = 1 \), but the overall estimation error can be reduced by assigning higher weightings to more reliable portions of data, provided that variation in reliability can to some extent be inferred. This section motivates a simple approach to weighting selection for WPS, and refines the estimate of effective weightings for EPZS.

The variance of a weighted average of independent measurements is minimised by setting weightings proportional to the reciprocal of the variance of each measurement. Point-wise displacement estimates in a window are not in fact independent, so ideally weightings should reflect the level of new information provided by each point. However, trying to predict the reciprocal of estimation variance still provides a suitable basis for considering possible weighting strategies:

\[
W(y) = \frac{\hat{\sigma}_u^2(y, \tilde{u}_k)}{\hat{\sigma}_u^2(y)}.
\] (4.6)

Five error sources (ES) affecting the variance of point-wise estimates following Equation 4.5a are listed below:

**ES1 Fine phase estimation error.** Displacement estimation error arises because \( \phi_{s1}(y) \neq \phi(y) \) and \( \phi_{s2}(y + \tilde{u}_k) \neq \phi(y_2) \), i.e., the common components within the recorded data are corrupted by noise components (cf., Equation 3.12). There will always be a degree of fine phase estimation error, with the variance of each phase estimate related to the ratio between common and noise component power.

**ES2 Bulk phase estimation error.** Ambiguity with regard to phase-wrapping is mostly avoided by the combination of robust tracking and using windows that are not so long as to have large differences between the displacements at either end. However, inevitably the wrapping of phase data is sometimes ambiguous. Incorrect wrapping of a phase value in any calculation produces a bulk phase estimation error of \( \pm 2n\pi \). It may be useful to assess the likelihood of bulk phase estimation error.

**ES3 Alignment error.** The approximation in Equation 3.20 leads to ambiguity regarding the accuracy of each point-wise displacement estimate if the pre- and post-deformation data are poorly aligned. Signal warping (e.g., adaptive stretching) is not incorporated as a component of the main deformation estimators in this thesis, owing to its relatively low computational efficiency. It follows that, if long windows
are used to analyse ultrasound data representing a high strain, points at either end of each window may be relatively poorly aligned, and consequently less reliable.

ES4 *Frequency estimation error.* Variation in \( \bar{\omega}(y) \) within a single window causes displacement estimation error. This would be an important factor to consider in the weighting strategy for ultrasound data of very high fractional bandwidth (e.g., > 100%). However, the significant aspect here is the frequency of the common component of the pre- and post-deformation signals, which is difficult to assess, rather than the frequency of the entire signal. Efforts to devise frequency-dependent weightings have not registered performance improvements working with the type of data used in this thesis.

ES5 *Tissue-signal displacement error.* The dominant displacement within the ultrasound signal sometimes fails to correspond to the displacement of the underlying tissue. This is usually due to limited resolution. A single, dominant scatterer can be the source of most of the ultrasound signal over an extended coherent-scattering region. Even if the underlying tissue is subject to a high strain, the displacement within the recorded signal will be equal to the displacement of the strong scatterer throughout the region that it dominates. More generally, it is not possible to resolve the displacements of multiple scatterers within a single resolution cell. These errors may be reduced in the future by improvements in ultrasonic resolution through progress with beamforming or deconvolution (topics beyond the scope of this thesis) [182]. The present analysis is aimed at accurately estimating the displacement present within the ultrasound signal.

**Signal amplitude**

A simple approximation can be derived for variance due to ES1 based on the properties of Equation 4.4, assuming that the common component power is larger than the noise power. An uncorrelated noise component of known power but unknown phase introduces equal variance in both the real and imaginary parts of the analytic signal, and zero covariance between them. Figure 4.1 shows the two signal components on an Argand diagram.

The noise contributes errors \( \Delta r \) to the real part and \( \Delta i \) to the imaginary part of the analytic signal. The common component power is assumed to be several times greater than the noise component power for most signal portions. Using the same notation as Figure 4.1, the phase error, \( \Delta \phi \), is approximated by

\[
\Delta \phi(y) \simeq \frac{p(y)}{f(y)} = \frac{\Delta r(y) \sin \phi(y) + \Delta i(y) \cos \phi(y)}{f(y)}
\]  

(4.7)
CHAPTER 4. PHASE-BASED ESTIMATORS

Figure 4.1: Fine phase estimation error: at a moderate ratio of common component power to noise component power, the phase estimation error, $\Delta \phi$, is inversely proportional to the common component envelope, $f$. Noise in the real and imaginary parts, $\Delta r$ and $\Delta i$, only translates into phase estimation noise through the component perpendicular to the common component, $p = \Delta r \sin \phi + \Delta i \cos \phi$.

For the purpose of weighting selection the estimation variance is considered:

$$
\sigma^2_\phi(y) = \mathbb{E} \left[ \Delta \phi(y)^2 \right] \quad (4.8a)
$$

$$
\simeq \mathbb{E} \left[ \left\{ \Delta r(y)^2 \sin^2 \phi(y) + 2 \Delta r(y) \Delta i(y) \sin \phi(y) \cos \phi(y) + \Delta i(y)^2 \cos^2 \phi(y) \right\} / f(y)^2 \right] \quad (4.8b)
$$

$$
\simeq \mathbb{E} \left[ \Delta r(y)^2 \right] \sin^2 \phi(y) + \mathbb{E} \left[ \Delta i(y)^2 \right] \cos^2 \phi(y) \quad (4.8c)
$$

$$
\simeq \frac{\sigma^2_n(y) \sin^2 \phi(y) + \sigma^2_n(y) \cos^2 \phi(y)}{\mathbb{E} \left[ f(y)^2 \right]} = \frac{\sigma^2_n(y)}{\mathbb{E} \left[ f(y)^2 \right]} \quad (4.8d)
$$

The product of uncorrelated real and imaginary errors in Equation 4.8b is zero under the statistical expectation operator. The expected squared errors, by contrast, are equal to the noise power. If the noise power is not estimated, then the predicted phase estimation variance is given by Equation 4.8d, expressing inverse proportion with the power of the common component.

Following Equation 4.6, the required variance is that of the point-wise displacement estimates. Inspection of Equation 4.5a shows that errors in the pre- and post-deformation phase estimates combine additively in the overall displacement error. For the purposes of devising a weighting, the unestimated noise power may be replaced by unity, leading to

$$
W(y) = \left( \sigma^2_\phi(y) + \sigma^2_\phi(y_2) \right)^{-1} = \left( \frac{\sigma^2_n(y)}{f(y)^2} + \frac{\sigma^2_n(y)}{f(y_2)^2} \right)^{-1},
$$

$$
= \frac{f(y)^2 f(y_2)^2}{\sigma^2_n(y) \left( f(y)^2 + f(y_2)^2 \right)} = \frac{f(y) f(y_2)}{\sigma^2_n(y) (c + c^{-1})} \quad (4.9)
$$


\[ c \] denotes the ratio of the common component envelopes, \( f(y_2)/f(y) \), which is unity by definition for perfect alignment, or otherwise close to unity. Therefore, neglecting variation in \( c \) and \( \sigma_n^2(y) \), a weighting based on the likely relative variance of fine phase estimation errors can be calculated if the product of common component envelopes is replaced by the product of the full envelopes of the recorded signals.

\[ W(y, u_k) = s_1(y)s_2(y + u_k) \]  

(4.10)

There is scope for improving this weighting strategy by relaxing assumptions made in this analysis, perhaps inferring the noise power. However, it is encouraging to note that weightings per Equation 4.10 resemble the approximation in Equation 3.24 for effective weightings with EPZS, since this is already known to have practical utility.

**Signal phase**

The weighting at each point can be scaled down depending on the phase separation, which is indicative of the likelihood of ES3 (bulk phase estimation error) and the level of ES4 (alignment error). The size of these effects is difficult to estimate, but the effective size of phase-dependent weightings in EPZS can be considered as a starting point.

For EPZS, the weighting approximation in Equation 3.22 indicates only amplitude and frequency contributions, following the small angle approximation in Equation 3.17. That might not be appropriate for long windows or high strains, in which case a better approximation is made by interpreting the scaling between phase and sine value as a phase-dependent weighting, so Equation 3.16 can be rewritten:

\[ \sum_{y=n\Delta y}^{n\Delta y+Y} W_A(y)W_B(y) \left( \phi(y_2) - \phi(y) \right) = 0, \]  

(4.11a)

where \( W_A(y) = f(y)f(y_2) \)  

(4.11b)

and \( W_B(y) = \frac{\sin\left(\phi(y_2) - \phi(y)\right)}{\phi(y_2) - \phi(y)} = \text{sinc}\left(\phi(y_2) - \phi(y)\right). \)  

(4.11c)

\( W_A \) is the amplitude contribution to the weighting approximation, and \( W_B \) is a phase weighting. Figure 4.2 illustrates the size of \( W_B \) for phase separations in the range \([-\pi, +\pi]\). This is interesting, partly because it may yield a better implementation of AMC for EPZS, and additionally because similar phase weighting can be applied in WPS.

The form of Equation 4.11 shows that, if phase values are wrapped to the same range, then WPS with appropriate selection of \( W_A \) and \( W_B \) can reproduce precisely the behaviour of EPZS. However, it is more interesting to experiment with alternative weightings to see if the level of displacement estimation error can be further reduced.

With regard to ES2 (bulk phase estimation error), it is immediately evident that, for a wrapping range of \([-\pi, +\pi]\), a value of phase separation equal to \( \pm\pi \) is meaningless, because an arbitrarily small error in the pre-wrapped phase leads to a knock-on error of \( \pm2\pi \). Values on the edge of the wrapping-range should be assigned zero weight, and values near
Figure 4.2: Plot of the phase contribution to weightings, $W_B(y)$, against phase separation, $\phi(y_2) - \phi(y)$, when using cross-correlation function phase.

the edge should be assigned low weights. This also penalises point-wise displacement estimates produced at relatively poor alignments (ES3). Particular phase weighting strategies probably can be motivated by minimum variance arguments for a particular statistical model, but a set of power-law heuristics is tested to gain initial insights:

$$W_B(y) = \left| \pi - \frac{\phi_{s2}(y + \hat{u}_k) - \phi_{s1}(y)}{\pi} \right|^h,$$

(4.12)

where the severity of the phase weighting is determined by $h$. Later tests span the range $h = 0$ (no phase weighting) up to $h = 3$ (heavy phase weighting).

4.3 Methods

4.3.1 Algorithms

The two families of deformation estimators are now compared. The implementation of EPZS has already been described in Section 3.3.1. The tests in Chapter 3 demonstrated that EPZS with robust tracking and AMC is a highly accurate deformation estimator compared with alternatives drawn from the literature. Various versions of WPS are now tested against variants of EPZS to compare performance, demonstrating a range of options that may be useful to algorithm designers.

WPS is another short-range search requiring initialisation at each window to within $\lambda/2$ of the correct displacement. Again, this is achieved by robust tracking methods from Chapter 2, from which cross-seeding is the crucial aspect. It requires an accuracy indicator for each displacement estimate, which need not be a correlation coefficient (see Appendix C) although in the test implementation correlation coefficients are calculated from scratch for each window after WPS has found a displacement estimate.

The full set of EPZS and WPS comparisons is listed in Table 4.1. The amplitude compression technique introduced in Chapter 3 is considered again, but rather than comparing different levels of log compression, only the limiting variant is retained, applying nonuniform gain to the signal to produce a uniform envelope, i.e., “discarded amplitude”, EPZS.L. Versions of EPZS are tested with and without AMC, also comparing the effect of the new weighting approximation following Equation 4.11, i.e., considering both amplitude and phase contributions to weightings. Calculation of the phase contribution is
cumbersome in EPZS, because point-wise phase data must be recovered by inverse tangent calculations, which are not required for any other part of the EPZS calculations (see Appendix C). Nonetheless, variants of EPZS with AMC considering the phase contribution to weightings (EPZS_A2 and EPZS_LA) are also tested.

The first stage of signal pre-processing for implementations of WPS is the same as for EPZS: real and imaginary parts of the analytic signal are produced by applying matched FIR filters to raw RF data (see Section 3.3.1). For WPS, the analytic signal is then converted to arrays of envelope and phase data following Equation 4.4, and phase is demodulated to baseband by subtracting $\omega_0 y$. The baseband conversion of phase data is useful, because it reduces the level of subsequent complexity involved in determining the correct wrapping of phase separations. The calculation, representation and manipulation of phase values may be performed very differently between naive and well-optimised implementations (see Appendix C).

Each iteration of the WPS search at each window is described by Equation 4.5c, starting with a seed displacement, $\tilde{u}_0$, and iterating until convergence. Subsample post-deformation data are calculated where necessary by linear interpolation of the baseband phase and envelope. To account for the demodulation, baseband phase separations are wrapped to $[\tilde{u}_k \omega_0 - \pi, \tilde{u}_k \omega_0 + \pi]$ to minimise phase-wrapping errors, and the iteration formula becomes:

$$\tilde{u}_{k+1} = \frac{\sum_{y=n \Delta y}^{n \Delta y+Y} W(y) \left(\phi_{s1}(y) - \phi_{s2}(y + \tilde{u}_k)\right)}{\omega_0 \sum_{y=n \Delta y}^{n \Delta y+Y} W(y)}.$$  \hspace{1cm} (4.13)

WPS variants are tested with and without amplitude weighting, phase weighting and AMC (see Table 4.1). Amplitude weightings following Equation 4.10, are the product of the pre- and post-deformation envelopes. Phase weighting basically follows Equation 4.12, now centred on the wrapping range $[\tilde{u}_k \omega_0 - \pi, \tilde{u}_k \omega_0 + \pi]$. Four levels of phase weighting are tested ($h = 0$ up to $h = 3$).

### 4.3.2 Tests

Performance is measured quantitatively using the same simulations as in Chapter 3 using $\text{SNR}_c$ as a performance measure (see Equation 3.4). Results are calculated using windows spaced at intervals of $2.7\lambda$ along each A-line, with strain calculated by the differencing method (see Equation 3.1).

A modified performance measure, $\text{SNR}_c^\beta$, is also calculated for some comparisons. One of the effects exposed by AMC is algorithm-dependent variation in the spacing between consecutive estimation locations. The mean location spacing (averaged over image area) is always greater than the window spacing, because each larger-than-average space covers a greater area than its counterpart smaller-than-average space. An increase in location spacing causes an increase in $\text{SNR}_c$, because the denominator in Equation 3.1 is larger, even if there is no effect on the precision of the displacement estimates. Increased location spacing is not in itself a positive consequence of changing parameters such as the weighting.
(a) EPZS variants.

<table>
<thead>
<tr>
<th>Amplitude</th>
<th>AMC</th>
<th>Off</th>
<th>On (env.)</th>
<th>On (env. &amp; phase)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>EPZS</td>
<td></td>
<td>EPZS_A1</td>
<td>EPZS_A2</td>
</tr>
<tr>
<td>Discarded</td>
<td>EPZS_L</td>
<td>-</td>
<td></td>
<td>EPZS_LA</td>
</tr>
</tbody>
</table>

(b) WPS variants.

<table>
<thead>
<tr>
<th>Amplitude</th>
<th>Phase weighting</th>
<th>AMC</th>
<th>Off</th>
<th>On</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>WPS_0</td>
<td></td>
<td>WPS_A0</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>WPS_A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>WPS_A2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>WPS_A3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discarded</td>
<td>0</td>
<td>WPS_L0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>WPS_LA1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Key to algorithms.

strategy, so $\text{SNR}_e^\beta$ can be calculated to remove that effect from the performance data:

$$\text{SNR}_e^\beta = \frac{\mu_e}{\sigma_e} \times \frac{\text{mean window spacing}}{\text{mean spacing of location estimates}}.$$  \hspace{1cm} (4.14)

This can be calculated to compare more sensibly between different algorithms that include AMC. On the other hand, it makes no sense to calculate $\text{SNR}_e^\beta$ if comparing between algorithms with and without AMC, because the mean spacing of locations is simply not known in the case without AMC. All of the improvement in $\text{SNR}_e$ associated with correcting the location estimates is valuable, and the improved location information can be doubly valuable where it avoids distorting the geometry of strain image features.

4.4 Results

Values of $\text{SNR}_e$ or $\text{SNR}_e^\beta$ are plotted against window length and strain. The quantitative results span a very wide range of window lengths, which helps with developing an understanding of the physical origins of performance differences, notwithstanding the fact that long windows undoubtedly reduce resolving power if they are used in practice.

Figure 4.3 compares the basic algorithms, EPZS and WPS_0. Performance without AMC and without AM suppression is of limited importance, since subsequent Figures 4.4–4.6 reinforce the need to handle amplitude modulation. Figure 4.7 illustrates the benefit of phase weighting. Results for the phase weightings by power-law heuristics are shown in Figure 4.8, which is the first test adopting $\text{SNR}_e^\beta$ as the performance measure.
Figure 4.3: Comparison of the basic algorithms, EPZS and WPS_0, tested at a range of strains and window lengths. They perform almost identically at 0.1% strain, while WPS_0 performs slightly better at 1% and 4%.

Figure 4.4: Comparison of the AMC variants applied to EPZS, tested at a range of strains and window lengths. Both implementations of AMC yield far higher performance than the basic algorithm. With long windows and high strains, the new correction in EPZS_A2 (accounting for phase weighting) outperforms the old correction.

substantial effect of phase weighting is to increase the spacing of the estimation locations, so comparisons based on SNR_e could be unfair.

Leading algorithms from the early results are compared in Figure 4.9. The main distinction is between discarded amplitude versus retained amplitude with AMC. Figure 4.10 provides a further comparison of SNR_e-strain characteristics. With short windows, these result in typical “strain filters” [81], but a different pattern emerges if the optimal window length is chosen for each algorithm at each strain.

Confusion could arise with the interpretation of some of these results, especially in cases of long windows and high strains. SNR_e data do not indicate whether performance degradation is caused by small numbers of outliers or by gradual degradation across the image. This is clarified by results in Figure 4.11, showing the same test as Figure 4.9 after the strain estimates have been median filtered.

The results up to this point are from simulations, which may fail to capture some aspects of real ultrasound scanning. Finally, to demonstrate that WPS works in conjunction with real data, Figure 4.12 shows in vitro and in vivo examples, where WPS_A1 has been applied to freehand scans using the hardware described in Section 1.4.
Figure 4.5: Comparison of EPZS.A2 against the discarded amplitude variants, tested at a range of strains and window lengths. Discarded amplitude algorithms mostly perform less well, but they are more robust when the windows are longer than optimal. The new AMC (accounting for phase weighting) improves performance in EPZS.LA.

Figure 4.6: Comparison of the WPS variants without phase weighting, tested at a range of strains and window lengths. The effects of AMC and discarded amplitude are similar to those observed with EPZS.

Figure 4.7: Comparison of the EPZS and WPS variants that retain amplitude and incorporate AMC, tested at a range of strains and window lengths. WPS.A0 performs less well overall than the other algorithms, because it omits phase weighting. This is important with long windows and high strains. Moderate phase weighting in WPS.A1 yields slightly different performance to EPZS.A2.
CHAPTER 4. PHASE-BASED ESTIMATORS

Figure 4.8: Comparison of WPS with different levels of phase weighting, $h$, at a strain of 0.5%. (a) SNR$_e$. (b) Mean spacing of location estimates. (c) Adjusted performance, SNR$_e^3$.

Figure 4.9: Comparison of leading algorithms recording SNR$_e^3$, tested at a range of strains and window lengths. (a) At 0.1% strain WPS_A0, WPS_A1 and EPZS_A2 have almost identical performance, with lower performance from the discarded amplitude algorithms. (b) At 1.0% strain the best performance comes from WPS_A1, followed by EPZS_A2, although discarded amplitude algorithms WPS_LA1 and EPZS_LA perform quite well and do better with long windows. (c) At 4.0% WPS_A0 performs marginally better than WPS_A1 and EPZS_A2 when the windows are short, but its performance drops away most quickly when the windows are long. The discarded amplitude variants mostly perform less well, although their robustness is advantageous.
Figure 4.10: Comparison of SNR\(_e\)-strain characteristics of leading algorithms, tested at various window lengths. The left and middle graphs show the shortest and longest window lengths at which results were recorded over all strains in the range 0.01–4%, while the right-hand graph shows the results for each algorithm where the optimal window length was employed at each strain. (Note that the discarded amplitude algorithm could not be tested at its optimum at low strains, because the range of window lengths permitted in the test software had an upper limit of 108.1\(\lambda\). Interpretation of the “best” results on the right is subtle, since window length is one of the factors that determines resolution.)

(a) SNR\(_e\) = 10.6 (b) SNR\(_e\) = 4.7

(c) SNR\(_e\) = 3.0 (d) SNR\(_e\) = 33.4

Figure 4.11: Effect of outliers. When within-window phase-wrapping errors begin to occur, initially there are only a handful of bulk errors. Wherever bulk errors occur, they register as extremely large strain errors, sufficiently large to skew the SNR\(_e\) value of the entire strain image, which exaggerates the effect on subjective image quality. For example, images are shown for EPZS\(_A2\) operating on a simulated strain of 4% with various window lengths (cf., Figure 4.4c). (a) 11.8\(\lambda\) ⇒ SNR\(_e\) = 10.6, (b) 12.7\(\lambda\) ⇒ SNR\(_e\) = 4.7, (c) 14.5\(\lambda\) ⇒ SNR\(_e\) = 3.0. (d) 14.5\(\lambda\) with outliers removed by a 3.5mm lateral median filter ⇒ SNR\(_e\) = 33.4. (e) Leading algorithms at 4.0% strain are compared in conjunction with median filtering (cf., Figure 4.9c).
CHAPTER 4. PHASE-BASED ESTIMATORS

Figure 4.12: Images from freehand strain imaging with WPS_A1 using a window length of 15λ. (a) B-scan and (b) strain image of breast biopsy phantom (0.42% relaxation). (c) B-scan and (d) strain image from biceps in vivo (0.75% relaxation).

4.5 Discussion

Generally, the results indicate that high-performance deformation estimation could be based on either WPS or EPZS. On the other hand, modifications between different variants bring potentially significant changes in performance, sometimes simply better/worse, but also more subtle differences, where an algorithm may be preferable under particular scan conditions.

The meaning and limitations of SNR_e as a performance measure are reconsidered before examining specific results. Arguably, the best feature of SNR_e is that, most of the time, it aligns closely with the level of noise that is perceived subjectively in “uniform” strain images. An obvious limitation is that effects on resolution need to be considered separately. Resolution is partly affected by window spacing, and displacement estimates of a given accuracy yield better strain estimates if they are differenced over wider spacing. This is one aspect of post-filtering, which is examined in Chapter 5. SNR_β is a preferable measure for comparing the absolute accuracy of displacement estimators that include AMC, as explained in Section 4.3.2. However, Chapter 5 will show that increases in window length also reduce resolution. When inspecting graphs of performance against window length, a tenfold increase in window length yielding a tenfold increase in SNR_e indicates much less than a tenfold increase in real performance, while an increase in window length without an increase in SNR_e indicates a reduction in performance for practical purposes.

Figure 4.3 compares the basic algorithms without AMC, for which the window length that maximises SNR_e is inversely related to the strain. The remarkable aspect of these results is that SNR_e actually varies little with changes in strain and window length. The deformation signal at 4% strain is 40 times greater than at 0.1% strain, but the peak value of SNR_e increases by a factor of just 1.5. This is primarily because uncorrected AM imposes a strain-independent performance limit. Additionally, higher strains cause signal decorrelation. WPS_0 slightly outperforms EPZS because it omits phase weighting, which amplifies location perturbations.
Figure 4.4 shows the improved performance in EPZS achieved by adding AMC, demonstrating that the modified weighting approximation in Equation 4.11 is more accurate. However, it is also interesting to note the window length beyond which SNR_e falls, which is here referred to as the drop-length. This is caused by the onset of a significant rate of large errors due to within-window phase-wrapping, which occurs when the strain multiplied by the window length is a substantial fraction of \( \lambda \). At a high strain, such as 4.0%, the drop-length is modest, hence it is an important consideration. Windows of around half the drop-length or above include regions where signal misalignment is substantial, even when the bulk displacement estimate is good, so the small angle approximation in Equation 3.17 no longer applies, but the new weightings from Equation 4.11 are significantly more accurate. Hence, these are also situations in which phase weighting matters, i.e., where EPZS_A2 achieves higher SNR_e than EPZS_A1.

Figure 4.5 shows that discarding the amplitude also improves performance compared to basic EPZS (it suppresses location perturbations), though less so than AMC, repeating a result from Chapter 3. Additionally, Figure 4.5 shows that the discarded amplitude variant of EPZS can be improved by the new form of AMC, considering phase weighting, since regions of substantial signal misalignment effectively have different weightings despite uniform amplitude. The advantage of discarded amplitude compared to variable amplitude is that estimation locations are nonetheless fixed nearer the centre of each window. This extends the drop-length by delaying the onset of within-window phase-wrapping errors.

Figure 4.6 shows that these principles extend to the WPS algorithms. These algorithms do not include phase weighting, so the results are affected more severely by the onset of within-window phase-wrapping errors, which makes the drop-lengths shorter. Figure 4.7 compares algorithms with and without phase weighting. SNR_e with short windows and at low strains is essentially unaffected by phase weighting, but it is important with longer windows and higher strains. The explicit application of phase weighting in WPS_A1 and the implicit phase weighting in EPZS_A2 extend the drop-length and boost performance compared to WPS_A0.

Figure 4.8 compares the effects of different forms of phase weighting. While Figure 4.8a suggests significant benefit from heavy phase weighting \((h = 3)\), Figure 4.8b shows that the results are distorted by changes in the mean location spacing. SNR^β_e is a better performance measure in Figure 4.8c. Different forms of phase weighting have quite similar impacts on the overall performance. Therefore, moderate phase weighting \((h = 1)\) is used for the remaining tests. Leading variants across all algorithm classes are compared under the adjusted performance measure in Figure 4.9. Variants with retained amplitude and AMC do outperform discarded amplitude variants. However, if discarded amplitude is selected, phase weighting (WPS LA1 and EPZS LA) significantly boosts performance compared to WPS_L0.

SNR^β_e-strain characteristics are compared for a smaller set of algorithms in Figure 4.10, showing estimation performance for a particular window length. The left and middle

---

\[ ^1 \text{Recall that ASE has no drop-length, so in some situations it outperforms algorithms that only incorporate AMC (see Figure 3.7).} \]
plots are typical “strain filters”, similar to characteristics that have appeared for other deformation estimators in previous literature [81]. The different algorithms with retained amplitude and AMC have similar performance. Phase weighting certainly lends robustness under some circumstances, but these plots demonstrate wide regions of behaviour in which phase weighting has little effect. The discarded amplitude example mostly has lower performance, but it performs best at high strain in the middle plot where the drop-lengths of other algorithms are exceeded. The SNR\(^{\beta}\)-strain characteristic for “best” window lengths, on the right, essentially indicates SNR\(^{\beta}\) just prior to the drop-length for each algorithm at each strain. This is not a normal strain filter, since long windows are associated with lower resolution. However, the negative gradient of the “best” plot above 1% strain is interesting, since high strains yield higher SNR\(^{\beta}\) for any given window length prior to the drop-length. The trend has practical significance, in that the drop-length is shown to represent no real constraint on the window length at low strains, but it becomes a serious issue at high strains.

While the drop-length is a salient property for characterising algorithm performance, its importance should not be exaggerated. The drop in SNR\(_e\) reflects a transition to different statistical behaviour above the drop-length. It does not really represent a point of catastrophic decline in estimation performance. Rather, displacement estimation error prior to the drop-length is characterised by a general level of estimation noise from a short-tailed distribution, which can be approximated as Gaussian. Increasing the window length reduces the variance of the distribution of errors, so SNR\(_e\) increases steadily. The significant event at the drop-length is the appearance of a second error mode related to bulk errors. The Gaussian component of the error distribution does not necessarily have increased variance, but ambiguous within-window phase-wrapping occasionally produces much larger errors. The outliers result in a much longer-tailed distribution for which SNR\(_e\) is an inappropriate performance measure (see Figure 4.11).

The occurrence of outliers is practically significant for the reasons discussed in Chapter 2. Some image processing techniques, such as compounding of image sequences, are highly appropriate in the presence of Gaussian noise, and behave poorly in the presence of outliers. The outliers can sometimes be removed. For example, median filtering removes errors in Figure 4.11d–e, so the drop-length is extended and SNR\(_e\) continues to rise. It would be wrong to assume that retained amplitude algorithms cannot be used in applications where the drop-length may sometimes be exceeded, but in those cases the post-processing needs to be designed to detect outliers. Equally, it is possible to envisage scanning tasks that could involve very high strains (probably not freehand quasistatic elasticity imaging) where the drop-length would be a severe limitation, regardless of non-linear post-processing. It might be possible to improve this behaviour by modifying the phase-wrapping within each WPS window. Otherwise, signal warping would be required.

Simulation results do not imply that the same behaviour will definitely be repeated in the context of real scanning, especially given that WPS is an unusual type of deformation estimator. Example images from freehand scanning in Figure 4.12 are provided simply to demonstrate that, as expected, WPS does work in practice. The targets shown are
a breast biopsy phantom (Computerised Imaging Reference Systems, Inc.\textsuperscript{†} Model 052) and human biceps \textit{in vivo}, showing strain estimated by a least squares filter spanning 15\(\lambda\) axially and 5 A-lines laterally. The combination of deformation estimation followed by least squares filtering is examined in detail in the next chapter.

Regarding the speed of these deformation estimators, suitable optimisation is discussed in Appendix \textsuperscript{C}. Images such as Figure \textsuperscript{4.12} are produced above 10 Hz (whether WPS or EPZS is used) during freehand scanning, using a single laptop with a 2 GHz processor running Stradwin software, receiving sampled RF data, performing all pre-processing and strain estimation for a live strain display.

### 4.6 Conclusion

WPS and EPZS both entail simplicity and low computational cost compared to most alternatives in the literature, and are suitable for real-time processing on single-processor computers. The form of WPS calculations is very different to EPZS, so it offers new options for algorithm optimisation, which is an ongoing research topic. Detecting signal envelope and signal phase is a major part of the computational cost in WPS variants, followed by a fast iterative search. The discarded amplitude variant with uniform weightings has the advantage of most computations being additions, although the multiplications in variants with nonuniform weightings usually substantially improve performance.

Developing the WPS approach has illuminated the properties of phase-based algorithms in general, and WPS offers a distinct advantage in terms of versatility. Current performance may be similar to EPZS, but WPS could be a good vehicle for exploiting progress in understanding the properties of RF ultrasound data from deformed scatterer fields. Relevant knowledge can be applied to improve the weighting strategy and investigate within-window phase-unwapping, which might extend the drop-length.

The analysis in this chapter only considered deformation estimation in the axial direction. It may be desirable to extend WPS to estimate displacements in the lateral and elevational directions by detecting lateral and elevational phases and envelopes. It has previously been shown that phase-based methods can be adapted for displacement estimation in the lateral direction \cite{183}. However, the higher axial sampling rate and ultrasound frequency is not the only reason for concentrating on this direction. The lateral direction is often under-sampled, certainly with the hardware used for this thesis \cite{184}. To achieve significant benefits from multi-dimensional displacement estimation beyond coarse lateral tracking is likely to require careful control of the lateral and elevational profiles of the point spread function, and dense sampling in the lateral and elevational directions. This is a beamforming problem, which falls beyond the scope of this thesis. Suitable lateral and elevational waveforms and sampling may or may not prove to be common features of future ultrasound machines.

\footnote{http://www.cirsinc.com}
Part III

Image formation
Chapter 5

Dynamic resolution selection

5.1 Background

As well as a high frame rate [80, 139, 185], freehand quasistatic elasticity imaging systems need to produce images that are robust when the size of deformation and the level of decorrelation are both variable [14, 139]. Good strain images can often be produced without dealing with this issue, but progress is needed for ease-of-use so that quasistatic elasticity imaging becomes a practical technique.

Accurate strain images cannot be produced from arbitrary pairs of pre- and post-deformation ultrasound frames. The frames must to some extent be correlated, which sometimes they may not be [14, 186]. Considering frame pairs where strain imaging is possible, the quality is still variable. The algorithms described in Part II are efficient methods for achieving accurate strain estimation based on given analysis parameters, but parameters yielding the best strain image vary from one frame to the next. This is affected by variable motion of the probe, leading to deformations of irregular sizes, and varying levels of decorrelation due to lateral and elevational movement. The suitability of particular parameters also depends on properties of the scan target, including the geometry and echogenicity of tissue regions with different mechanical features.

In general, strain estimation precision can be improved by analysing at coarse resolution. Mean strains can be estimated easily over large distances, whereas fine-scale displacement gradients are more problematic. An image with greater detail is only more informative if the detail has acceptable SNRe. For image interpretation, clinical experience is likely to indicate the relative importance of sharp definition versus accurate, quantitative measurement of strain contrast.

Constructing a strain image from pre- and post-deformation ultrasound frames can be considered in two stages, both involving several parameter choices. Deformation estimation discussed in Part II requires choices of window length, width, and spacing. To form a strain image the displacement field is then differentiated, which can be performed by piecewise-linear least squares regression (PLLSR) [163]. Accuracy and resolution are affected by the length and width of the least squares kernel. In past investigations of strain imaging the analysis parameters have been fixed to particular settings, although
CHAPTER 5. DYNAMIC RESOLUTION SELECTION

experienced researchers know how to modify the image properties by modifying the settings [33].

The first aim in this chapter is to examine the effects of strain imaging parameter selection on resolution and SNR. Secondly, that information is applied in a system for dynamic resolution selection (DRS), whereby parameters are steered automatically towards good settings for each image. This is motivated by two desirable outcomes. DRS should reduce the complexity presented to the sonographer, reducing the control options to a single noise-rejection setting, which deals appropriately with inter-image variability. Additionally, the parameters may vary within each image to produce reliable data despite intra-image variability in the levels of strain and decorrelation [164, 166].

This principle is pursued at a conceptual level, and is implemented for the case of WPS from Chapter 4 with AMC from Chapter 3 followed by PLLSR. Theory and empirical evidence from analysis of simulation and phantom data are considered while developing and validating DRS, which is also demonstrated under in vivo conditions.

5.2 Development

The parameters associated with both deformation estimation and strain estimation incur some reduction in resolving power, because deformation fields cannot be sampled with acceptable accuracy at the level of individual samples in RF ultrasound signals. The dimensions of the analysis units are important. In this chapter, “windows” always refers to displacement estimation, and “kernels” refers to strain estimation by PLLSR. The size of the windows and kernels governs a tradeoff between estimation precision and resolution. SNR defined in Equation 3.4 is evaluated to assess accuracy. SNR can always be improved by sacrificing resolving power in the strain images. The following analysis first considers the effect of parameter selection on resolution. Second, the influence of strain estimation parameter values on the translation from displacement estimation errors into strain estimation error is examined. Third, the original source of noise is considered, i.e., displacement estimation error, focusing on the role of displacement estimation parameters in noise reduction. Finally, these concepts are combined practically in Section 5.2.5, forming the basis for DRS.

5.2.1 Resolution

The most obvious restriction on strain image resolution comes from the spacing between neighbouring estimation locations, but it is less important than filtering effects related to the size of windows and kernels. The only disadvantage of dense estimation spacing is increased computation time, while it improves both resolution and SNR. The estimation locations used in this chapter are very closely spaced, so as to focus on the significance of the other parameter choices.

Accuracy is increased by estimating strain at lower resolution, but the ultrasound scanner should preferably have very high resolution, because this reduces the level of decorrelation.
CHAPTER 5. DYNAMIC RESOLUTION SELECTION

Figure 5.1: Effect of a MA filter on the contrast between low and high strain bands. (a) The axial resolving limit is the axial length scale at which features are just resolved, with a strain field that is uniform in the lateral direction. (b) Displacement and strain fields against distance. (c) A MA filter (length=$L_r$) easily resolves the different strain regions. (d) Resolution is still achieved with length=$1.5L_r$. (e) The contrast is zero when the filter length is $2L_r$. (f) Filter lengths $>2L_r$ register negative contrast.

The effects on strain estimates of changes to window and kernel size are similar, though not identical, to denoising by applying a moving average (MA) filter.\textsuperscript{†} The overall error decreases as a MA filter gets larger, because uncorrelated error components spanned by the filter average to zero. However, the output from a large MA filter has coarse resolution.

There is no universal definition of resolution that can be applied sensibly to all imaging tasks. For the present analysis, the resolving limit is defined as the feature scale, $L_r$, at which there is no longer any positive contrast between two or more bands of low strain sandwiching and surrounded by background material with higher strain (see Figure 5.1). The resolving limit is reached when strain estimates after the filter exhibit zero contrast between the low and high strain bands. For example, the resolving limit of a MA filter is half the filter length.

The resolving limit is assumed similarly to be approximately proportional to window and kernel dimensions, although the constant of proportionality may not be the same in both cases. This is investigated empirically in Section 5.3.3. The greatest estimation accuracy at a given resolution will usually be achieved by windows and kernels of the

\textsuperscript{†}The analogy becomes inaccurate for window size approaching the drop-length (see Chapter 4). This issue receives further attention in Section 5.2.4.
5.2.2 Strain estimation error

Differentiation amplifies estimation noise, particularly if the displacement estimates are closely spaced. Figure 5.2 illustrates this point for 1D strain estimation. Predicting strain estimation error requires an understanding of how displacement estimation error filters through.

There are various techniques for reducing noise in gradient estimates, including low-pass filtering and wavelet denoising. The following analysis focuses on piecewise-linear least squares regression, which is more commonly applied. Analysis of alternative approaches would involve similar considerations.

The simplest unweighted least squares gradient estimate is

\[
\hat{\varepsilon}_m = \frac{\sum_{\Omega_n \in \mathcal{K}_m} \tilde{y}_n \hat{u}_n}{\sum_{\Omega_n \in \mathcal{K}_m} \tilde{y}_n^2},
\]

where \(y\) denotes axial distance. The strain estimate, \(\hat{\varepsilon}_m\), is produced using data from a set of displacement estimation windows, \(\{\Omega_n\}\), comprising displacement estimates \(\{\hat{u}_n\}\) at locations \(\{\tilde{x}_n, \tilde{y}_n\}\) (measured relative to the centre of kernel \(\mathcal{K}_m\)). Note that, although \(\{\tilde{y}_n\}\) are axial location estimates, the kernel is usually 2D, taking displacement estimates from multiple neighbouring columns. The set of displacement estimate locations must be symmetric about the centre of the kernel to ensure that the strain estimate has greatest validity at that point.\(^1\)

The scale of strain estimation errors is predicted by evaluating the variance of this estimator. In general, there are errors in both displacement and location, all of which

---

\(^1\)Asymmetric kernels estimate with greatest validity away from the kernel centre, which can distort the geometry of image features, although location estimation similar to AMC from Chapter 3 can easily be derived for least squares kernels.
lead to strain estimation error. Location errors in \( \tilde{y}_n \) are assumed to be negligible, having been substantially reduced by AMC (see Chapter 3). This leaves errors only in \( \hat{u}_n \), resulting in the following strain estimation variance if covariances are negligible:

\[
\sigma^2_{\varepsilon_m} = \frac{\sum_{\Omega_i \in K_m} \tilde{y}_i^2 \sigma^2_{u_i}}{\left(\sum_{\Omega_i \in K_m} \tilde{y}_i^2\right)^2}.
\] (5.2)

It is also possible that the displacement estimator introduces significant covariances, in which case a more complicated expression must be evaluated:

\[
\sigma^2_{\varepsilon_m} = \frac{\sum_{\Omega_i \in K_m} \sum_{\Omega_j \in K_m} \tilde{y}_i \tilde{y}_j \sigma_{u_i u_j}}{\left(\sum_{\Omega_i \in K_m} \tilde{y}_i^2\right)^2}.
\] (5.3)

Both approaches are tested in Section 5.3.4.

### 5.2.3 Displacement estimation error

#### Signal model

A suitable signal model is needed for analysing the displacement estimator to estimate errors and covariances for Equations 5.2 and 5.3. It is helpful to separate the signal into common and noise components as in Equation 3.12. In this chapter, displacement estimation is performed purely in the axial direction, without coarse lateral tracking. The common components are transformed by an axial displacement field identical to the axial displacement in the underlying tissue. The rest of the signal comprises noise components arising because of changes in speckle interference patterns brought about by changed scatterer spacing, decorrelation due to off-axis motion and electrical noise.

The signal model is re-expressed in Equation 5.4, now including the lateral index, \( x \), for analysing 2D windows. Subscripts 1 and 2 denote pre- and post-deformation signals. \( a_1 \) and \( a_2 \) are the pre- and post-deformation analytic signals, with envelopes \( s_1 \) and \( s_2 \) and phases \( \phi_{a_1} \) and \( \phi_{a_2} \) respectively. The axial displacement field in the tissue is \( u(x, y) \). Common signal \( f(x, y)e^{j\phi(x, y)} \) is warped precisely in accordance with the displacement field, so it is reproduced identically in Equations 5.4b and 5.4d, whereas the noise components, \( n_1(x, y)e^{j\phi_{n_1}(x, y)} \) and \( n_2(x, y)e^{j\phi_{n_2}(x, y)} \), are uncorrelated.

\[
a_1(x, y) = s_1(x, y)e^{j\phi_{a_1}(x, y)} = \text{common signal + noise signal}
\] (5.4a)

\[
a_2(x, y + u(x, y)) = s_2(x, y + u(x, y))e^{j\phi_{a_2}(x, y + u(x, y))} = f(x, y)e^{j\phi(x, y)} + n_1(x, y)e^{j\phi_{n_1}(x, y)}
\] (5.4b)

\[
a_2(x, y + u(x, y)) = s_2(x, y + u(x, y))e^{j\phi_{a_2}(x, y + u(x, y))} = f(x, y)e^{j\phi(x, y)} + n_2(x, y)e^{j\phi_{n_2}(x, y)}
\] (5.4c)

#### Displacement estimator

While it is common to evaluate theoretical lower bounds on the displacement estimation error [143, 144, 153, 155, 156], in this chapter a mechanism is sought for predicting the
actual size of error from a particular displacement estimator depending on properties of the recorded signals. The example considered is WPS from Chapter 4. The iterative search using baseband phase data was expressed in Equation 4.13, and is now re-expressed indicating the use of 2D windows. The formula for the \( k \)th iteration at the \( n \)th window, \( \Omega_n \), is as follows, where \( \omega_0 \) is the assumed probe centre frequency used for demodulating to baseband:

\[
\tilde{u}_{k+1} = \frac{\sum_{\{x,y\}\in\Omega_n} W(x, y, \tilde{u}_k) \left( \phi_s1(x, y) - \phi_s2(x, y + \tilde{u}_k) \right)}{\omega_0 \sum_{\{x,y\}\in\Omega_n} W(x, y, \tilde{u}_k)},
\]

(5.5a)

where

\[
W(x, y, \tilde{u}_k) = W_A(x, y, \tilde{u}_k)W_B(x, y, \tilde{u}_k),
\]

(5.5b)

\[
W_A(x, y, \tilde{u}_k) = s_1(x, y)s_2(x, y + \tilde{u}_k),
\]

(5.5c)

and

\[
W_B(x, y, \tilde{u}_k) = \pi - |\phi_s1(x, y) - \phi_s2(x, y + \tilde{u}_k)|/\pi.
\]

(5.5d)

Iterations converge on the displacement estimate, \( \hat{u}_n \). Then AMC is applied to minimise location errors, which for 2D windows gives both lateral and axial co-ordinates:

\[
\hat{x}_n = \frac{\sum_{\{x,y\}\in\Omega_n} W(x, y, \hat{u}_n)x}{\sum_{\{x,y\}\in\Omega_n} W(x, y, \hat{u}_n)},
\]

(5.6a)

\[
\hat{y}_n = \frac{\sum_{\{x,y\}\in\Omega_n} W(x, y, \hat{u}_n)y}{\sum_{\{x,y\}\in\Omega_n} W(x, y, \hat{u}_n)}.
\]

(5.6b)

In Chapter 3 it was observed that regular window positions result in an irregular distribution of estimation locations. At some point the array of estimates must be returned to a regular grid for display purposes. This can be performed by interpolation. However, in the current chapter WPS is tested with a modification whereby the 2D windows are spaced irregularly so as to target regular estimation locations in both the axial and lateral directions. With 1D windows this entails accumulating the location error while samples are added to the window, taking each next sample from the end that opposes the current location error, as mentioned in Section 3.4.5. The algorithm modification for 2D windows of a fixed aspect ratio is less simple, but works by extending the same principle.

**Error prediction**

First consider the errors in point-wise displacement estimates (PWDEs). The concept of estimating displacement at a particular sample was introduced in Chapter 4. Section 4.2.2 included a theoretical derivation considering fine-scale phase estimation error (assuming correct phase-wrapping) when the common component of the signals is stronger than the noise component, leading to Equation 4.9, showing that PWDE variance is proportional

\[\text{This estimator is WPS_A1 in the notation of Chapter 4.}\]

\[\text{These location estimates are recorded in image co-ordinates. The expression of least squares strain estimation in Equation 5.1 requires subtracting the kernel centre to obtain the window locations in kernel co-ordinates, i.e., } \{\tilde{x}_n, \tilde{y}_n\}.\]
to the square of the noise power divided by the square of the envelope of the common
signal. This motivates a modified definition of the ultrasonic signal-to-noise ratio,

$$\text{SNR}_s(x, y) = \frac{f(x, y)^2}{\frac{1}{2}(n_1(x, y)^2 + n_2(x, y)^2)}.$$  \hspace{1cm} (5.7)

The variance of displacement estimates produced by WPS is expected to be near propor-
tional to $\text{SNR}_s$. Furthermore, when a set of uncorrelated PWDEs is averaged, constituting
a window, the displacement estimation variance is expected to scale inversely with the
number of samples.

In practice, not all PWDEs are mutually independent, so displacement estimation
variance is greater by a factor proportional to the autocorrelation length of the errors.
The average degree of correlatedness between PWDEs depends on window length. How-
ever, once windows are longer than the error autocorrelation length, the addition of each
new sample adds an almost equal amount to the total information. This means that dis-
placement estimation precision usually becomes a near-linear function of window length,
as for example in Figure 5.3a. PWDEs in neighbouring A-lines are only weakly correlated,
so estimation precision versus window width is usually more linear (see Figure 5.3b). It
is therefore reasonable to predict displacement estimation precision with the product of
$\text{SNR}_s$, window size and a constant of proportionality, $K$. Window dimensions are denoted
by $X$ and $Y$ in the lateral and axial directions.

$$\sigma_u^2 \simeq (K XY \text{SNR}_s)^{-1}$$ \hspace{1cm} (5.8)

An estimate of $\text{SNR}_s$ is required so as to make use of this result. If the underlying tissue
deformation, $u(x, y)$, were known exactly, $\text{SNR}_s$ could be estimated with high accuracy
as a function of the real part of the arithmetic correlation coefficient of correctly warped
pre- and post-deformation analytic signals. The real part of the arithmetic correlation
coefficient of windows aligned using displacement estimates is expressed as

$$\rho(\Omega_n, \hat{u}_n) = \frac{\sum_{\{x,y\} \in \Omega_n} \Re\{a_1^*(x, y)a_2(x, y + \hat{u}_n)\}}{\frac{1}{2} \sum_{\{x,y\} \in \Omega_n} |a_1(x, y)|^2 + |a_2(x, y + \hat{u}_n)|^2}.$$ \hspace{1cm} (5.9)

For perfectly aligned data, the terms of the product of pre- and post-deformation samples are

$$a_1^*(x, y)a_2(x, y + u(x, y)) = f(x, y)^2 \left[ + f(x, y) \left( n_1(x, y)e^{j(\phi(x, y) - \phi_{n1}(x, y))} + n_2(x, y)e^{j((\phi_{n2}(x, y) - \phi(x, y))} \right) 
+ n_1(x, y)n_2(x, y)e^{j((\phi_{n2}(x, y) - \phi_{n1}(x, y))} \right]. \hspace{1cm} (5.10)$$

The product terms are all uncorrelated, except for the square of the common envelope.
Similarly, if the denominator of Equation (5.9) is expanded for perfectly aligned data, the
only correlated terms are the sum of the squared common envelope and half the sum of
squared noise envelopes. If they are integrated over a reasonable number of samples, the
CHAPTER 5. DYNAMIC Resolution SELECTION

113

Figure 5.3: Displacement estimation precision versus window dimensions for simulation data with a mean SNR of 20 dB subject to a uniform compression of 0.01%. The horizontal scale is in samples of data with a 6.0 MHz centre frequency sampled at 66.7 MHz, so 500 samples correspond to 5.75 mm. 12 A-lines correspond to 3.75 mm. (a) 1D windows of lengths in the range 10–500 samples. The dotted line is included so that the curved section can be identified easily. (b) Analysis with windows of length 120 samples and widths in the range 1–12 A-lines.

uncorrelated terms tend to zero, so an estimate of SNR_s is produced by applying the following transformation.

\[ \hat{\text{SNR}}_s(\Omega_n, \hat{u}_n) = \frac{\rho(\Omega_n, \hat{u}_n)}{1 - \rho(\Omega_n, \hat{u}_n)} \quad (5.11) \]

The problem in practice is that the true displacement, \( u(x, y) \), is unknown; \( \hat{u}_n \) is only approximately correct. Nevertheless, SNR_s can be roughly estimated by assuming that \( \hat{u}_n \) is near-correct at all points in window \( \Omega_n \); Equation 5.11 potentially provides a useful estimate, with some restrictions.

Two types of alignment error limit the accuracy of \( \hat{\text{SNR}}_s \). (1) Displacement estimation error (error in \( \hat{u}_n \)) leads to misalignment throughout the window, which can lead to over- or under-estimation of SNR_s. (2) Intra-window misalignment caused by intra-window strain is more problematic, introducing severe bias into \( \hat{\text{SNR}}_s \), because it always leads to under-estimates. Figure 5.4 shows examples at different strains. Displacement estimation error is the main source of error at 0.01% strain, so values of \( \hat{\text{SNR}}_s \) from longer windows are slightly more accurate. The situation is reversed at higher strain, where intra-window misalignment becomes the main source of error. Figure 5.4b shows the effect on SNR_s at different window lengths. The disadvantage of long windows is that the range of \( \hat{\text{SNR}}_s \) values becomes much narrower than the true range. Estimates from shorter windows are more consistent across different strains. SNR_s values from long windows could be improved by signal warping, but shorter windows are preferable when this is not pursued.

Accurate values of \( \hat{\text{SNR}}_s \) are not in themselves sufficient for predicting strain estima-
Figure 5.4: SNR$_e$ estimates at different 1D window lengths for simulation data similar to that used in the previous figure, with various levels of electrical noise. The vertical axis is image-wide SNR$_e$, which has been evaluated by averaging the arithmetic correlation coefficients across all windows and substituting the result into Equation 5.11. Data are plotted for two levels of uniform strain. (a) 0.01%. (b) 0.5%.

5.2.4 Importance of the strain level

In Section 4.5 displacement estimation error due to phase-wrapping at the edges of long windows was noted to be mainly an issue at high strains. With increasing window length, SNR$_e$ performance flattens as point-wise phase-wrapping errors become more common. Beyond a certain drop-length, which is inversely related to the strain, there is a severe reduction in performance. SNR$_e$ contour plots in Figure 5.6 illustrate this effect. At a low strain phase-wrapping errors never occur, so SNR$_e$ goes up with every increase in window and kernel length. At higher strains, as in Figure 5.6b, SNR$_e$ no longer rises so quickly beyond a certain window length, and eventually it drops. The relationship between window length and estimation precision is also less linear when there is phase-wrapping (see plots in Chapter 4), so error prediction may be inaccurate. As a rule-of-thumb, the window length should not be selected greater than \[\frac{5.0}{\text{strain magnitude}}\] samples$^\dagger$ to stay in the linear region, considering a 66.7 MHz sampling rate with a 6.0 MHz centre frequency. For example, if there is a minimum window length, say 50 samples, then the maximum strain that can reasonably be measured is 10%.

The strain level is also important because it determines the scale of significance of

$^\dagger$This equals \((0.45/\text{strain magnitude})\lambda\).
CHAPTER 5. DYNAMIC RESOLUTION SELECTION

Figure 5.5: Examples of normalised covariance of displacement estimation between nearby windows from analysis of simulation data at 20 dB with a uniform compression of 0.01%.
(a) Normalised covariance against axial separation. Displacement estimates separated in the axial direction become uncorrelated less quickly if they are from long windows.
(b) Reciprocal of axial decorrelation rate against window length. The dotted line aids visual assessment of the divergence from linearity.
(c) Normalised covariance against lateral separation. Displacement estimates separated in the lateral direction remain correlated longer when the windows are wide.
(d) Reciprocal of lateral decorrelation rate against window width. Both axially and laterally, the separation of windows at which the covariance tends to zero is approximately equal to the window dimension.

strain estimation errors. The analysis so far has considered predicting the absolute size of strain estimation errors. Performance in terms of SNR also depends on the size of the strain signal (see Equation 3.4). The noise level in a homogeneous image should obviously be measured against the mean strain, but the appropriate benchmark for inhomogeneous images is less obvious. Large regions of mostly low strain may be swamped by noise if the significance of errors is held to depend on the image-wide mean strain. However, if local strain sets the scale, then zero strain regions (which do arise in practice) always
Figures 5.6: Effect of phase wrapping (the drop-length) on displacement and strain estimation performance with simulation data at an SNR of 20 dB in the presence of electrical noise, with further decorrelation caused by a 0.0625mm elevational translation of the scatterer field. For the horizontal axis, kernel length in samples comes from the length in windows multiplied by the spacing between windows. (a) SNR<sub>e</sub> contours against 1D parameter settings at 0.5% strain. (b) SNR<sub>e</sub> contours at 2.0% strain.

register zero SNR<sub>e</sub>. A compromise is mainly to consider local strain values, saturating when they deviate wildly from the mean. This is the approach adopted for DRS, but all three options are compared qualitatively in Section 5.3.6.

### 5.2.5 Dynamic Resolution Selection procedure

The above strands of analysis are brought together in DRS to improve the value of real strain images. The tasks in DRS divide between three stages illustrated with a flow chart in Figure 5.7. The aim is to set the parameters automatically so that a threshold performance, minSNR<sub>e</sub> (set by the user), is always exceeded.

**Survey.** An initial survey produces displacement and correlation data using WPS with short windows throughout the image. The search algorithm includes cross-seeding, multipass analysis and continuity checking from Chapter 2. Short windows are used for two reasons. Firstly, the windows must be sufficiently short to operate without severe phase-wrapping throughout the stated dynamic range. A length of 50 implies that strains should not exceed 10%. Secondly, short windows are preferred because the resulting SNR<sub>e</sub> values exhibit relatively little variation in bias across different strains. 1D survey windows could be used for fast processing, but wide 2D windows (say 10 A-lines) may offer greater robustness. Both widths are tested in Section 5.3.4. The displacements are converted to strain estimates, ū, using a very large least squares kernel of 51 windows (510 samples) by 17 A-lines. Additionally, the arithmetic correlation coefficients are smoothed with a spatial Gaussian filter; the smoothed values are transformed into signal-to-noise estimates,
SNR_s. The subsequent Analysis requires SNR_s and \( \hat{\varepsilon} \) values throughout the image, so Survey data are extended to image edges by nearest-neighbour extrapolation.

**Analysis.** At each estimation location a search is undertaken to find the window and kernel dimensions of finest resolution that should produce performance exceeding \( \text{minSNR}_e \). Trial values for window and kernel dimensions are combined with \( \text{SNR}_s \) and \( \hat{\varepsilon} \) values from the Survey to obtain \( \text{SNR}_e \). Steps required include the prediction of displacement estimation error using \( \text{SNR}_s \) and window size (Equation 5.8), estimation of normalised covariance between windows in a least squares kernel by evaluating their overlap fractions, and the substitution of both sets of data into the formula for strain estimation error. This can be evaluated with Equation 5.2 if covariances have little impact, or (more rigorously) with Equation 5.3 if covariances are significant. This is converted into an SNR_e predic-
CHAPTER 5. DYNAMIC RESOLUTION SELECTION

Figure 5.8: Artefact caused by large window and kernel changes over a short distance. (a) Region from DRS strain image at boundary between soft background and stiff inclusion phantom. The boundary appears twice because of excessive window and kernel overlap. (b) Window lengths (white=500 samples). (c) Window lengths have been clipped, thus eliminating the double-boundary artefact.

...tion by substituting the strain estimation error along with an appropriate indicator of the strain level into Equation 3.4. The strain level is usually taken to be the local estimate, $\hat{\varepsilon}$, but extreme values saturate at half and twice the mean strain.

Each time $\hat{\text{SNR}}_e$ is calculated, the search terminates if it exceeds $\min \hat{\text{SNR}}_e$. Otherwise, the window and kernel dimensions are increased for a new set of trial parameters. Kernel length is increased in increments of two windows to maintain symmetry about a particular estimation location. Kernel width also increases according to the aspect value. Window length and width are increased as well, in an appropriate ratio determined in Section 5.3.3. However, the window length is subject to a limit of $[5.0/\text{local strain}]$ to avoid phase-wrapping, while window width continues to grow after the length has saturated.

Eventually adequate window and kernel dimensions are found to exceed $\min \hat{\text{SNR}}_e$ at every estimation location (or they saturate at maximum values). One final analysis step is required before proceeding with Refinement. The parameter selections are parsed to eliminate an artefact that arises if parameter values change wildly over small distances. If any window or kernel extends further in any direction than its neighbour on that side, this can result in the appearance of a higher level of information than is actually present. For example, a monotonic strain gradient could appear non-monotonic. This is eliminated trivially by clipping window and kernel lengths, so they never extend beyond 100% overlap with a neighbour. Figure 5.8 shows an example.

Refinement. Displacement and strain estimation is repeated using the DRS parameter values. Rather than robust tracking following Chapter 2, the seed for each first iteration of WPS is taken from the Survey window with the highest arithmetic correlation coefficient in the same row, within five columns of the active estimation location.

5.3 Experiments

Experiments were undertaken to determine values of unknown constants required in DRS, and to assess the behaviour of the resulting system. All results were produced with
estimation locations distributed over a regular grid, with lateral spacing equal to the spacing between A-lines, and axial spacing equal to 10 samples (slightly less than one cycle at the centre frequency). The windows and kernels were 2D, with a fixed aspect ratio such that the length in samples was 30 times greater than the width in A-lines.

5.3.1 Image formats

Various types of image are presented: displacement estimates, strain estimates, DRS window lengths and local SNR. They have a consistent format. Data are obtained at every estimation location before being displayed, and the spacing is dense enough for linear interpolation to produce an acceptably smooth display.

A main motivation for this investigation is to examine performance variation at different resolutions. Therefore, a block pattern covers data at image edges, where estimates could not be evaluated using the stated parameter settings. Displacement and window length images are blocked-out at estimation locations where the windows did not fit inside the recorded data frame. Blocked-out borders on the strain images are roughly twice as thick, since strain data is displayed only from kernels that were filled entirely with valid windows. SNR images have the same borders as the corresponding strain images. For practical scanning the borders would be unnecessary, although it must be accepted that SNR/resolution cannot be controlled arbitrarily at edges.

The images are normalised consistently. Values that determine the image scales are indicated in the figure captions. All valid displacement data are first processed in a single least squares filter to estimate the mean strain. This sets the mid-grey level in the strain images, with saturation at zero (black) and twice the mean strain (white). Similarly, displacement images have a mid-grey level at the mean strain multiplied by half the image height, with saturation at zero (black) and twice this value (white). SNR images have mid-grey set to the image-wide SNR, with saturation again at zero and twice this value. Finally, window length images are scaled such that the maximum window length is displayed white, mid-grey is half that value and black would be a window of zero length.

5.3.2 Simulations

Simulations using Field II [147] were performed to model the ultrasound system detailed in Section 1.4. Simulations were generated to test a range of different properties. The main simulation parameters are the same as described in Section 3.3.2. Further details are listed in Table 5.1.

5.3.3 Resolution

The effects of window and kernel size on resolution were tested by producing strain images using a range of 1D windows (30–1000 samples) and 1D kernels (10–1000 samples), recording the contrast (difference of mean strains) between the low and high strain bands
### Table 5.1: Details of simulations.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Reference(s)</th>
<th>Scattering field</th>
<th>Other details</th>
</tr>
</thead>
<tbody>
<tr>
<td>resolution experiment</td>
<td>Simulation A</td>
<td>$2.0 \times 10^5$ scatterers, equal scattering strengths, 50$\times$50$\times$6 mm volume (lateral$\times$axial$\times$elevational), uniform spatial distribution</td>
<td>background strain 1.0%, zero strain bands either side of focus, various feature scales, (cf., Figure 5.1)</td>
</tr>
<tr>
<td>effect of electrical and uncoloured noise</td>
<td>Simulation B-20dB, Simulation B-15dB, etc.</td>
<td>as per Simulation A</td>
<td>range of levels of white noise, mean SNR is appended to simulation reference</td>
</tr>
<tr>
<td>effect of lateral motion decorrelation</td>
<td>Simulation C-0.0625mm, Simulation C-0.125mm, etc.</td>
<td>$3.0 \times 10^5$ scatterers, uniform dist. scat. strengths, 51$\times$50$\times$8 mm volume, uniform spatial distribution</td>
<td>range of lateral motions, motion size appended to simulation reference</td>
</tr>
<tr>
<td>effect of elevational motion decorrelation</td>
<td>Simulation D-0.0625mm, Simulation D-0.125mm</td>
<td>as per Simulation C</td>
<td>elev. motion appended to simulation reference</td>
</tr>
<tr>
<td>performance with nonuniform noise</td>
<td>Simulation E</td>
<td>$1.0 \times 10^6$ scatterers, equal scattering strengths, 50$\times$50$\times$6 mm volume, spatial distribution is nonuniform, gets more dense left-to-right</td>
<td>mean SNR 20dB, but SNR varies throughout the image, uniform strain 1.0%</td>
</tr>
</tbody>
</table>

at the vertical centre of Simulation A. The thresholds at which the contrast goes from positive to negative indicate filter combinations at the resolving limit. Thresholds from four different feature scales are combined in Figure 5.9f, to form a contour map of resolving limit.

The contours are mostly rectangular, indicating that the resolution is limited essentially by either window or kernel length, whichever has the greater smoothing effect, so it is reasonable to set the window and kernel lengths in a fixed ratio to achieve optimal estimation performance for any particular resolving limit, since increasing the window or kernel size almost invariably reduces estimation noise (cf., Figure 5.6). The extremes on the resolution contours are plotted in Figure 5.9f. These data are close to relationships of direct proportion between filter length and resolving limit. Least squares estimates of the constants of proportionality indicate that the kernel length should be 1.20 times greater than the window length for any particular resolution. This ratio is used for parameter selection in the Analysis stage of DRS, as described in Section 5.2.5.
CHAPTER 5. DYNAMIC RESOLUTION SELECTION

(a) contr. = 0.781  (b) 0.492  (c) 0.229  (d) 0.0600  (e) -0.0357

Figure 5.9: Strain estimation resolution assessed by analysis of Simulation A. (a–e) Example strain images (mean strain 0.85%) of a 2.5 mm feature scale viewed with various parameter settings. Contrast is indicated as a fraction of the ideal contrast, ranging from high contrast in the first image to failure to resolve in the final image. The window/kernel lengths were 75/90, 175/210, 275/330, 375/450 and 525/630 respectively, using 2D windows and kernels. 500 samples corresponds to a physical length of 5.75 mm. (f–g) Quantitative results with 1D windows and kernels. (f) Contour plot of resolving limit over parameter setting space. (g) Maximum parameter values from each contour plotted against feature scale.

5.3.4 Calibration

Implementing DRS requires a scale factor, $\sqrt{K}$, for Equation 5.8. This section presents calibration results. Simulation data were processed with a range of uniform window and kernel settings (2D kernels in the range 70–610 samples; 2D windows 1.20 times shorter). These were performed instead of the Refinement stage in DRS, preceded by the Survey (windows of length 50 samples, either 1 or 10 A-lines wide). Rather than adjusting the window and kernel settings, Analysis produced a squared error prediction for each strain estimate. Strain divided by the root mean squared error prediction is image-wide $\hat{SNR}_e$ (before scaling by $\sqrt{K}$), which can be compared to the measured value, $SNR_e$. A suitable value for $\sqrt{K}$ was found by comparing the predictions and measurements across a range of different strains, window/kernel dimensions and noise conditions.

By way of example, Figure 5.10a shows $SNR_e$ against $\hat{SNR}_e$ for Simulation B-20dB at 0.5% strain. This is far from the expected straight line because of baseline error in the
Figure 5.10: DRS calibration. (a) Measured SNR$_e$ against SNR$_e$ predictions across a wide range of parameter settings, without correction for baseline error. (b) Temporal shift rate of the PSF centroid against scatterer depth. This nonlinearity is a source of baseline error. (c–f) The effect of baseline error has been removed prior to evaluating the remaining results. (c) Calibration curves for Simulation B-20dB, if overlap covariances are ignored (Equation 5.2). The fit to any linear relation is poor. (d) Calibration curves for Simulation B-20dB with covariances considered (Equation 5.3). The correlation coefficient between predictions and measurements is 0.999, with a scale factor, $\sqrt{K} = 0.186$. (e) Calibration results with covariances considered from various simulations for a 50×1 survey. (f) Calibration results for a 50×10 survey.
strain estimates: The apparent depth in ultrasound images (time of receiving echo) does not vary perfectly linearly with scatterer depth (see Figure 5.10b). This occurs in real scans too, which are subject to further baseline error caused by fluctuations in the speed of sound. Baseline error sets a ceiling of $\text{SNR}_e \approx 50$ in these simulation results. So as to avoid degrading the estimate of $\sqrt{K}$, the calibration is based on measured $\text{SNR}_e$ values not greater than 30, from which the baseline error is removed as follows, assuming a baseline $\text{SNR}_e$ of 50:

$$\text{corrected SNR}_e = \left( \text{measured SNR}_e^2 - \text{baseline SNR}_e^2 \right)^{-\frac{1}{2}}. \quad (5.12)$$

Corrected $\text{SNR}_e$ measurements are plotted against $\text{SNR}_e$ predictions in Figure 5.10c for a particular scatterer field, subject to various strains. Covariances are ignored, so the calculation follows Equation 5.2. This is shown to be inapplicable, since the relation between predicted and measured $\text{SNR}_e$ is neither linear within any single curve, nor do the curves at different strains come from a single function. Covariance estimates based on overlap have been used with Equation 5.3 to compute better $\text{SNR}_e$ predictions in Figure 5.10d. This shows a much higher degree of linearity, with similar curves at the different strain levels. The least squares gradient (estimate of $\sqrt{K}$) is 0.186, with a correlation coefficient of 0.999 between the predictions and the measurements.

The type of noise may affect the value of $\sqrt{K}$, so calibration data from different types of simulations are shown in Figure 5.10e–f. These combine results across strains of 0.01%, 0.5%, 1.0% and 5.0%. The white noise in Simulation B can be thought of as the extreme opposite noise category compared to motion decorrelation in Simulations C and D. $\text{SNR}_e$ drops in the motion simulations because of a noise ultrasound signal with similar statistical properties to the common signal. Despite this difference, within each simulation the linear relationship between predicted and measured $\text{SNR}_e$ is good, evidenced by the high correlation coefficients. The difference between the $\sqrt{K}$ values of the different noise categories is discussed in Section 5.4. The variation means $\text{SNR}_e$ cannot be controlled precisely by DRS, but it may nonetheless be useful, since $\sqrt{K}$ varies by much less than an order of magnitude. The mean calibration value is $\sqrt{K} = 0.146$ for the 50×1 Survey or 0.131 for the 50×10 Survey. The remainder of the experimental results use 50×10 Surveys.

### 5.3.5 Variable noise experiment

Simulation E has a variable noise level. Lateral variation in SNR is indicated by the mean SNR throughout different A-lines, shown in Figure 5.11a. Within each A-line, the SNR is highest at the focus, so the Survey produces a range of correlation coefficients illustrated in Figure 5.11b. A normal strain image is noisier in the low correlation regions (Figure 5.11c), but the variation is reduced by applying DRS (Figure 5.11d).

DRS should maintain fixed $\text{SNR}_e$ throughout the image. A performance measure, $\mathcal{V}$, records spatial variation in $\text{SNR}_e$. This is evaluated by Gaussian filtering\(^1\) of squared

\(^{1}\)The Gaussian filter has the same dimensions as the filter for correlation that was mentioned in Section 5.2.5.
Figure 5.11: Variable noise simulation (1.0% strain). (a) Mean SNR values by A-line. (b) Smoothed arithmetic correlation coefficients from the Survey (black=0.7, white=1.0). (c) Strain image with fixed parameters, image-wide SNR<sub>e</sub>=7.70 and V=2.08. (d) Strain image with DRS, SNR<sub>e</sub>=7.51 and V=4.63. (e) Window length (max=91), strain and SNR<sub>e</sub> images for fixed parameter strain imaging, image-wide SNR<sub>e</sub>=11.2 and V=2.15. (f) Window length (max=175), strain and SNR<sub>e</sub> for DRS, SNR<sub>e</sub>=12.87 and V=5.24. (g) V against SNR<sub>e</sub> both for DRS and for fixed parameter values. Higher V indicates that DRS reduces the variability of SNR<sub>e</sub>. 
Figure 5.12: Homogeneous phantom subject to 3.6% compression. (a) Survey strain. (b) DRS window length. (c–d) Strain and local SNR for 1: fixed parameters and 2: DRS parameters.

errors in strain to obtain local estimates of SNR. These are displayed as images in Figure 5.11e–f. $V$ is calculated by dividing the overall SNR by the standard deviation of local SNR estimates. $V$ is improved substantially by applying DRS instead of fixed parameters.

5.3.6  **In vitro and in vivo scanning**

Data were recorded during scans using the system described in Section 1.4. Strain images were produced by analysis of consecutive frames of acquired data, applying DRS in offline processing. One subject is a homogeneous phantom. This is an agar cylinder (height 50mm, diameter 69mm) with 0.6wt% Al$_2$O$_3$ powder providing suitable scattering. A footprint extender was attached to the probe to apply a uniform compression. This is an interesting application for DRS, especially at high strains with a 1D estimator, since some regions decorrelate due to lateral expansion. A compression of 3.6% has been applied in Figure 5.12. Strain images in Figures 5.12c1–c2 exhibit more uniform performance when DRS is applied. Figure 5.12b shows that larger windows and kernels are required at the sides, where lateral Poisson displacement is greatest. The $V$ values vary by only a small amount, perhaps because the baseline error in assuming uniform strain is quite high in this case.

To apply DRS to *in vitro* frames with varied SNR, the data in Figure 5.13 are from the freehand elevational sweep of Figure 1.17 for freehand 3D strain imaging. The elevational movement introduces noise, and in some instances the strain signal is very small.
Figure 5.13: Illustration of different strategies for consecutive frames from freehand 3D scanning of an olive-gelatin phantom. Mean strains are noted on the right, where negative values indicate relaxation. (a) Fixed parameter settings for relatively fine resolution. (b) DRS with minSNR_e=5.0. (c) Fixed parameter settings for coarser resolution.
Figure 5.14: Demonstration of strain selection for the SNR_e calculation, using scan of human biceps in vivo (image mean strain of -0.65%, minSNR_e=6.0). (a) Survey strains. (b) Window length and strain images when SNR_e is based on the mean strain across the image. Low strain regions may be swamped by noise. (c) Images basing SNR_e on the local survey strain. Low strain regions are analysed at extremely low resolution. (d) Compromise where SNR_e is based on local strain with saturation.

Figure 5.13 compares DRS, with suitable minSNR_e, against two fixed-parameter settings, one coarser than the other. DRS is shown to combine the advantages of both fixed settings, not smoothing unnecessarily when the data quality is good, but going to a coarse resolution in regions of low SNR_e to avoid presenting meaningless data.

Further frames are shown from a freehand scan of biceps in vivo. Figure 5.14 illustrates the strain benchmark issue (cf., Section 5.2.4). The extrapolated Survey strains are shown in Figure 5.14a. The SNR_e calculation based on mean strain leads to the DRS
CHAPTER 5. DYNAMIC RESOLUTION SELECTION

Figure 5.15: Range of choices for $\text{min} \hat{\text{SNR}}_e$ in another frame from the scan of human biceps (image mean strain of -0.64%).

strain image in Figure 5.14b. Basing SNR$_e$ on the local strain, as in Figure 5.14c, gives a smoother image in the low strain region, but the window settings cover a very wide range. Figure 5.14d shows the image using the normal compromise (local strains with saturation). Finally, Figure 5.15 shows another image from the same scan, processed at a range of different $\text{min SNR}_e$ levels. Qualitative changes in smoothness and variability are obvious, although clinical experience may be required to determine appropriate settings for specific tasks.

5.4 Discussion

The calculation of SNR$_e$ developed in Section 5.2 included numerous approximations, particularly the assumption of overlap fraction corresponding to the normalised covariance between neighbouring windows. However, the high degree of linearity in Figure 5.10d and the close fit to the linear model indicated by correlation values in Figure 5.10e–f serve as partial validation of this approach.

The results also highlight a limitation: The calibration factors vary by as much as a factor of two, separating into two groups based on whether decorrelation is primarily due to white noise and strain decorrelation or to non-axial motion. In fact, this is to be expected. Equation 5.8 relies on the error autocorrelation distance having a constant value (cf., justification of Equation 5.8 in Section 5.2.3), which determines $\sqrt{K}$. Errors with long-distance correlations cause greater loss of performance than short-distance errors with equal SNR$_e$. A more accurate prediction could be made if the error autocorrelation distance could be estimated. Otherwise, the approach followed here was to choose a suitable, typical value for $\sqrt{K}$, which offers useful if imprecise control over SNR$_e$.

The variable noise experiment showed how DRS can respond to within-image variation in decorrelation. This is evident from comparison of the SNR$_e$ images in Figure 5.11e–f, supported by the uniformity data in Figure 5.11g. At low SNR$_e$, DRS is similar to fixed-parameter analysis, because it simply opts for the minimum window and kernel dimensions if minSNR$_e$ is already exceeded throughout the image. At moderate SNR$_e$, however,
window and kernel dimensions are closely controlled so that $\text{SNR}_e$ is roughly uniform, as illustrated in the plot of $\mathcal{V}$ against $\text{SNR}_e$. $\mathcal{V}$ becomes a less reliable measure at high $\text{SNR}_e$, however, since baseline errors eventually distort the results. Overall, this is strong evidence that DRS can lead to greater uniformity in $\text{SNR}_e$ by means of varying the resolution.

However, it may not be intuitively obvious that uniform $\text{SNR}_e$ is a sensible goal. As compared to the DRS image (Figure 5.11f), while the fixed parameter image (Figure 5.11e) is noisier in the corners, it clearly has higher $\text{SNR}_e$ at the focal depth and on the right of the image. The important property of DRS is that $\text{minSNR}_e$ sets the degree of significance of contrast differences, so the information that is presented can be interpreted meaningfully. Sometimes this may blur over subtle features below the resolution setting, but the resolution level is at least discernable to the user, since a ripple of magnitude proportional to $\text{SNR}_e$ is present with a wavelength that depends on the imaging parameters. By contrast, an image with variable noise at fixed resolution presents data where the significance of image features is less well defined, and interpretation is highly dependent on a process of cognitive filtering; users must discriminate for themselves between those features that are meaningful, and others that are assumed to be noise.

The homogeneous phantom results in Figure 5.12 are further evidence that DRS can simplify image interpretation. The variable noise due to off-axis motion is a typical strain imaging phenomenon associated with high strains. No claim is made here that 1D tracking is optimal for high-strain work, where the addition of lateral tracking becomes advantageous. Against that, the outcome in Figure 5.12c–d2 is encouraging, because it shows that DRS can adjust for the decorrelation caused by off-axis motion, so as to maintain the reliability of the image data. Indeed, this result is generally useful, because lateral motion increases signal decorrelation even if lateral tracking is employed.

Figure 5.14 illustrates a complication to this issue, in that $\text{SNR}_e$ is ill-defined in the context of inhomogeneous scan subjects. Should noise be measured against the local strain, the local level of strain variation (perhaps a strain envelope) or simply compared to the same mean strain throughout the image? The mean strain option is unsatisfactory, because if a substantial region has relatively low strain, the local information can be completely lost under a noise level that is acceptable elsewhere in the image. On the other hand, it is obvious that the local strain cannot be used for $\text{SNR}_e$, because strain zero crossings result in points where $\text{minSNR}_e$ can never be exceeded, even at the coarsest resolution (e.g., towards the bottom of Figure 5.14e). The compromise of saturated local strain may be sensible, though there is undoubtedly scope for assessing more rigorously the effect on information content of estimation noise in different parts of the image.

Nevertheless, in the present form DRS already behaves preferably compared to fixed-parameter settings when applied to real data. The sequence of images from a freehand 3D sweep in Figure 5.13 shows that DRS moves between resolution settings in a manner that appears intuitively reasonable, so images are neither overly smoothed where the signal is strong, nor are misleading estimates displayed at inappropriately high resolution in regions of poor data. The question of exactly what level is suitable for the $\text{minSNR}_e$ setting
remains open, although various settings are illustrated in Figure 5.15. The optimal setting may depend on the degree to which users can discriminate for themselves between noise and signal features in poor image regions. For example, new users might benefit from a high minSNR to obtain reliable data (sometimes at low resolution). Experienced users may be able to extract meaning from noisier images at finer resolution by comparing the image data with their expectations. The significance of noise may also be application-dependent.

5.5 Conclusion

Resolution and estimation error have been examined in detail for a particular strain estimation algorithm. It was argued that the best noise performance at any resolution may be achieved by window and kernel dimensions set in a fixed ratio. A viable technique was outlined for predicting displacement estimation error, and translating it into strain estimation error.

These observations have been applied in DRS, whereby data are automatically processed with suitable parameters to achieve a desired SNR whilst allowing the resolution to vary. This produces strain images in which the significance of image features can be interpreted with greater confidence.

A key task that has not been addressed is adjusting DRS to reduce the computational load. The present system is time-consuming. The implementation of WPS for DRS in this chapter omitted optimisation techniques such as those described in Appendix C, and was somewhat slower owing to the modification to estimate displacement at regular locations. Even fairly basic implementations of WPS can usually produce at least one 2D strain image per second. For DRS, with no attempt at optimisation, each strain image takes at least 30 seconds on a current standard PC with a 3 GHz processor, and extremely poor data takes up to five minutes if DRS opts for long, wide windows. The algorithm is suited to a high level of parallel processing, operating simultaneously on each A-line, so real-time DRS processing could perhaps be achieved even with DRS in its current form. However, the algorithm would need substantial modification to achieve real-time imaging using standard, single-processor equipment. The Analysis is not the issue. Displacement estimation at both the Survey and Refinement stages can vary hugely in cost depending on numerous factors. Major costs are incurred by the unusually large number of estimation locations in the present version, and the occasional use of very large 2D windows. Several thousand samples are often covered by a single window. However, this approach can perhaps be combined with intelligent downsampling at various stages to reduce the computational load whilst avoiding excessive performance penalties.

DRS-like results may be achievable by using a fixed window size with better optimised filtering for strain estimation. This falls beyond the scope of this thesis, although it has been tackled in other investigations by the author, extending the analysis in this chapter [170, 171]. The alternative approach has numerous interesting properties including much lower computational cost.
Chapter 6

Practical 2D and 3D interface

6.1 Background

In this chapter an alternative approach is considered to providing a clinical interface for freehand quasistatic elasticity imaging. Given the highly interactive nature of ultrasound examinations, established scanning modes have advantages in that clinicians are already well practised in the required scanning techniques, understand the significance of typical images, and are generally familiar with the uses, benefits and disadvantages of each mode. Sonographers have extensive experience with conventional ultrasound imaging (B-mode greyscale), colour Doppler and power Doppler [27]. The likelihood of an addition to the ultrasound tool-set gaining clinical favour may be boosted if it possesses an interface that is practically helpful: actively fostering the development of a successful scanning technique, by providing visual or audio feedback; displaying data in an intuitively meaningful format; and automatically guarding against the presentation of misleading data.

Those issues concern how information is presented. It is also necessary to decide what information to present. This point was raised in Section 1.4. Qualitatively, what type of information can be provided (stiffness, strain, or a compromise)? Quantitatively, how much data should be amalgamated to form each display image? This question is highly relevant to 2D and 3D elasticity imaging using freehand ultrasound. Persistence (temporal filtering) may help to improve the live display of a continuous sequence of 2D images. Spatial averaging may improve the display of 3D data spanning a volume.

Regarding the type of information, this chapter describes pseudo-strain imaging in detail. Undoubtedly, strain images can be misleading, because an interpretation of low strain as indicating high stiffness becomes erroneous when the stress field varies substantially [13, 14, 33, 88] (cf., Section 1.4). It would be preferable to derive Young’s modulus images if possible by solving the inverse problem, but theoretical and practical difficulties with doing so were discussed in Sections 1.3.1 and 1.4. Nevertheless, the concept of a level of applied stress is necessary so that strain can be related to mechanical properties of the tissue. Some patterns of variation in the stress field occur repeatedly, and can be adjusted for. Strain normalisation with a scale-factor that varies both between images and within every image is a constrained, heuristic approach to reducing ambiguity, producing pseudo-strain images.
In practice, the basic challenge of achieving an acceptable strain estimation signal-to-noise ratio remains a more immediate obstacle to freehand strain imaging, despite the developments of Chapters 2-4. Many frames produce good images, but a substantial fraction may be difficult to interpret because of high estimation noise. DRS from Chapter 5 may be a good way to extract information from individual images, but important features may often be missed when images are produced at low resolution. Another common approach to noise reduction using multiple images is to perform averaging over a sequence [187]. This chapter presents an adaptive weighted approach, which can be used for persistence in live displays of 2D images, and spatial filtering in displays of volumetric data.

Novel normalisation and weighted filtering methods are aspects of a practical interface for 2D and 3D pseudo-strain imaging, which builds on the algorithm developments of earlier chapters. The aim in this chapter is to describe the interface concept in general terms; the reader may envisage numerous specific applications. The interface is designed considering the demands of freehand scanning, but the principles may have wider relevance. The key aspects should be applicable to enhancing any quasistatic strain imaging system, almost regardless of the approach taken in the earlier stages of signal processing, although some restrictions are discussed. The rationale behind each aspect of the interface is demonstrated using example images from freehand scans of in vitro and in vivo targets.

6.2 Method

Illustrations in Sections 6.3 and 6.4 are provided based on the example of WPS with AMC,† with the full set of tracking strategies from Chapter 2 (cross-seeding, multi-pass analysis, continuity checking, secure initialisation and coarse lateral tracking), except where otherwise stated, followed by strain estimation by PLLSR [163]. This offers a good demonstration, primarily because it has already been tested rigorously, especially in Chapter 5, resulting in a promising method for predicting strain estimation variance, and additionally because the accuracy of this approach is good in the context of the strain imaging literature. Further details of the sources of data and specific parameter settings are provided alongside the results in Sections 6.3 and 6.4.

However, the interface methods are presented in general terms where possible. This section begins with an overview of the interface as a whole, followed by a brief discussion of predicting estimation precision, and descriptions of the three subsequent interface processing stages: normalisation, persistence or spatial filtering, and display (see Figure 6.1).

†Specifically, the version used here is the optimised implementation of WPS described in Appendix C, including amplitude and phase contributions to the weightings.
6.2.1 Interface concept

The suitability of ultrasound data for producing strain images varies substantially depending on scanning technique, physiological motion and changes in the analysis parameters. Although the parameters can be controlled as in Chapter 5, adjusting to an extent for different properties, this cannot overcome all of the difficulties associated with practical strain imaging. At some stage it becomes impossible to produce meaningful deformation data from frames that are extremely weakly correlated. Depending on scanning technique, an adequate minimum level of correlation may not always arise. With a very poor technique, it may not even occur often. In the majority of frames where uniform SNR can be achieved by adjusting the resolution settings, it is desirable to improve the quality of the recorded ultrasound data to achieve the maximum resolving power. The best data may arise from relatively substantial deformations (i.e., typically a large fraction of 1%, sometimes lower or higher depending on the target) accompanied by relatively low decorrelation. The acquisition of good data depends on combined properties of the scanning technique and the tissue.

Strain imaging with a typical interface requires a high level of expertise both in scanning technique and image interpretation [36]. Figure 6.2 illustrates common difficulties. A side-by-side display showing B-scans next to strain images has been suggested for practical elasticity imaging [33], because it is then easier to match strain data with features of interest in the B-scan. Sometimes there may be little or no data, owing to an absence of coupling to the tissue. A side-by-side display helps in this case exemplified by Figure 6.2a–b because the sonographer knows when to ignore the strain display. Image interpretation may nonetheless be rather difficult when the coupling is good, as in Figure 6.2c–d, because some scan targets do not provide signals suitable for strain imaging throughout the entire image.

One approach that to some extent handles this problem is a display of strain data overlaid on the B-scan as a “colour wash”, where colour indicates strain and brightness is partly determined by the ultrasound signal amplitude [34, 35, 40]. Ultrasound signal amplitude correlates with the accuracy of strain estimates, so this goes some way to indicating the quality of the strain data. It only helps to a limited extent, however, since signal amplitude is a very weak indicator of overall decorrelation. While a complete absence of signal would certainly mean that strain estimates were dominated by noise, it is often the case that strain estimates from regions with medium signal amplitude are less noisy than other estimates where the signal is stronger. Furthermore, the blend of strain with B-mode data could actually make insightful image interpretation more difficult, by mixing strain data with fine features of B-scans such as the speckle pattern, that are not...
generally related to tissue stiffness.

The approach in this chapter is based on the availability of more accurate indicators of the precision of each strain estimate, which can be applied to influence the use of these estimates at every stage downstream including the display. An appropriate strain normalisation can be calculated by fitting a constrained surface to the entire set of displacement data in each frame. Normalisation can be applied both to the strain data and to the associated precision data, producing a new array of pseudo-strain data with updated precision values. Having produced a single frame of pseudo-strain, the signal-to-noise ratio can be boosted by applying some form of persistence or spatial filtering, which again can be weighted by precision. The output is a set of persisted pseudo-strain values and appropriately updated precisions. Finally, the display scheme can be tailored to indicate both strain and precision on a 2D scale represented by a 2D colour map.

### 6.2.2 Predicting estimation precision

This system exploits the availability of useful predictions of strain estimation precision. The interface is likely to be most advantageous if the precision predictions are highly accurate, but the form of the prediction method is not critical. The method employed here is based on the findings of Chapter 5.

To summarise, precision is the reciprocal of variance or mean squared error. Displacement precision can be predicted by evaluating $\frac{XYc}{(1 - c)}$, where $c$ is the correlation coefficient (with arithmetic normalisation) between pre- and post-deformation data in the
displacement estimation window, $Y$ is the window length and $X$ is window width. In each PLLSR kernel, the overall strain estimation variance can be predicted by evaluating an average of the displacement variances weighted by the square of the distance from the kernel centre, dividing by the sum of squared distances. The results in Sections 6.3 and 6.4 use the following approximation for strain estimation precision, $W_A$:

$$W_A(x, y) = \frac{(\sum_i \hat{y}_i^2)^2}{\sum_j \hat{y}_j^2(1 - c_j)/(X_j Y_j c_j)}$$  \hspace{2cm} (6.1)$$

where the sums are over displacement estimation windows in the PLLSR kernel centred on pixel $(x, y)$, and $\hat{y}$ denotes distance from the centre of the kernel along the axial direction in which strain is being estimated.

Chapter 5 showed that estimates of strain estimation precision can be produced with greater accuracy by applying a more complicated formula accounting for the correlations between nearby errors. This may improve results, but the simple formula above is applied now, because the computation is cheaper, and seems to produce good results in terms of relative precision when the imaging parameters (window and kernel dimensions) are fixed. Figure 5.4 in Chapter 5 also showed that correlation coefficients lead to biased estimates of displacement precision if the windows are very long. The windows used in Sections 6.3 and 6.4 have moderate length, for which the bias is not expected to be excessive. This compromise is adequate for illustrative purposes.

### 6.2.3 Normalisation

Careful design of the normalisation strategy may contribute to valuable improvements in the quality of pseudo-strain images, particularly if real-time images are required or the scanning procedure is freehand. Various approaches have been reported in the past [14, 33, 186]. The basic problem of finding an appropriate strain scale for each image can be solved robustly by fitting a plane to the entire set of displacement estimates, $\{u(x, y)\}$. This is performed in later examples by the method of precision-weighted least squares, thereby determining an average strain. The equation of the fitted plane is

$$\hat{u}(x, y) = \alpha + \hat{\epsilon}y.$$  \hspace{2cm} (6.2)$$

The strain estimates can be scaled so that the dynamic range in the display goes from zero up to a fixed multiple of the average strain, $\hat{\epsilon}$.

For the new interface, this approach is extended by fitting other parametric surfaces to the displacement estimates. For instance, the axial stress is typically lower at greater depth away from the probe, because the stress spreads out into the surrounding tissue. This can be adjusted for by fitting the following normalisation surface:

$$\hat{u}(x, y) = \alpha_0 + \alpha_1 x + \beta_1 y + \beta_2 y^2,$$  \hspace{2cm} (6.3a)$$

$$\hat{\epsilon}(x, y) = \beta_1 + 2\beta_2 y.$$  \hspace{2cm} (6.3b)$$
It is more appropriate to refer to \( \hat{\varepsilon}(x, y) \) as a “normalisation” strain, rather than an average, since it is a function of image position. The parameters \( \beta_1 \) and \( \beta_2 \) are again evaluated in later examples by precision-weighted least squares regression. Normalisation entails each strain estimate being divided through by the local value of \( \hat{\varepsilon}(x, y) \).

A further extension can be made to adjust for the probe rotating about the elevational axis during the scan, resulting in stress variation over the lateral direction:

\[
\hat{u}(x, y) = \alpha_0 + \alpha_1 x + \beta_1 y(1 + \beta_2 y)(1 + \beta_3 x), \quad (6.4a)
\]
\[
\hat{\varepsilon}(x, y) = \beta_1 (1 + 2\beta_2 y)(1 + \beta_3 x). \quad (6.4b)
\]

Again, \( \beta_1, \beta_2 \) and \( \beta_3 \) can be found by precision-weighted least squares regression, or any suitable alternative, defining the normalisation strain throughout the image. No doubt many other extensions are possible.

Whichever type of normalisation surface is used, the normalisation applies to the precision values as well as to the strain estimates. Normalisation scales strain measurements and errors by \( 1/\hat{\varepsilon}(x, y) \), so the precision (reciprocal of mean squared error) must be scaled by \( \hat{\varepsilon}(x, y)^2 \). If pre-normalisation strain estimates and post-normalisation pseudo-strain are denoted \( \varepsilon_A \) and \( \varepsilon_B \), with \( W_A \) and \( W_B \) denoting the pre- and post-normalisation precisions, then

\[
\varepsilon_B(x, y) = \varepsilon_A(x, y)/\hat{\varepsilon}(x, y) \quad (6.5a)
\]
\[
\text{and} \quad W_B(x, y) = W_A(x, y) \times \hat{\varepsilon}(x, y)^2. \quad (6.5b)
\]

The practical effect of the combined normalisation is to place each individual strain estimate on a broad scale of possible interpretations, depending on its relative properties in the context of the entire frame of scan data. Depending on the value of \( \hat{\varepsilon}(x, y) \), the normalisation locates any single strain estimate and its precision value within a range spanning (1) relatively low pseudo-strain at relatively high precision, through to (2) relatively high pseudo-strain at relatively low precision. The form of the normalisation potentially influences not only the type of image, but also the level of noise.

It bears noting that the main computational expense of normalisation comes from fitting a parametric displacement surface, but this is typically a negligible cost on widely available GHz processors in the context of 2D frame rates below 100 Hz. Computational efficiency is one of the main factors behind the selection of the particular parametric forms that are provided here as examples. However, Equations 6.3 and 6.4 imply linear variation with depth in the normalisation strain. This may usually be a good approximation, but it leaves open the possibility that the sign of the normalisation strain could invert within the image. If this were a reasonable form of normalisation, it would imply that at some depth the direction of the stress field inverts, i.e., that a compression at the surface may cause extension at greater depth within the tissue. This is unrealistic, but it can be prevented, for example by constraining the fitted surface to avoid the strain crossing zero within the image depth (an approach taken in Sections 6.3 and 6.4) or alternatively data below the zero crossing could be treated as uninformative by setting the precision to zero.
Figure 6.3: Illustration of the types of strain fields that may produce uniform pseudo-strain fields — indicating homogeneous stiffness — with each of the normalisation options. Here low strain is shown by white and high strain is black. The first option only adjusts for stress variation on the level of the whole image, while the second adjusts for lower stress away from the probe surface, and the third also adjusts for uneven probe pressure.

The normalisation surface might ideally reflect exponential variation with depth, but the least squares fit would incur much greater computational cost. The normalisation surfaces outlined above are demonstrated as efficient options, not precluding the possibility that other parametric or constrained non-parametric forms may be found in the future, offering better performance at reasonable cost.

Figure 6.3 illustrates the strain fields that are implied by each of the normalisation schemes, or equivalently the stress fields that might produce such a field in homogeneous material. The key with the normalisation is to fit a suitably constrained surface, that with high probability corrects for artefacts associated with the uneven distribution of stress within the tissue, without removing information that has arisen owing to genuine differences in stiffness.

It is possible, if unlikely, that there may be tissue in which stiffness in fact varies with the reciprocal of depth, and the application of a uniform stress field may also be possible, in which case normalisation using Equation 6.3 or Equation 6.4 could remove real stiffness data from the display. The frequency with which this sort of ambiguity arises will depend on the scanning target, so the appropriate normalisation surface might depend on the clinical application.

6.2.4 Persistence or spatial averaging

Persistence refers to averaging image data over time, while in freehand 3D scanning it is more appropriate to apply spatial averaging. The concept of weighted averaging of multiple frames is not new in ultrasonic strain imaging [188]. In general, averaging after normalisation weighted by precisions on a per-pixel basis is expected to be the best approach. The precision values are also summed, since it can be shown that the overall precision of a correctly precision-weighted average in which the errors are uncorrelated is equal to the sum of precisions.
In the context of producing a live 2D display during freehand imaging, persistence is applied on the arrival of each new frame, $f$. The values that persist in pre-display buffers at pixel $(x, y)$ are a precision-weighted sum, $S(x, y, f)$, and the sum of precisions, $\Omega(x, y, f)$. The buffers are updated when each new frame arrives, providing new pseudo-strain data, $\epsilon_B(x, y, f)$, and new precision data, $W_B(x, y, f)$.

$$S(x, y, f) = \gamma S(x, y, f - 1) + W_B(x, y, f)\epsilon_B(x, y, f), \quad (6.6a)$$
$$\Omega(x, y, f) = \gamma \Omega(x, y, f - 1) + W_B(x, y, f). \quad (6.6b)$$

$\gamma$ is a number greater than 0.0 and less than 1.0 that determines the level of persistence. Each persisted pseudo-strain is given by $S(x, y, f)/\Omega(x, y, f)$, accompanied by a precision (quality) value for the display, $\Omega(x, y, f)$, so sonographers can be presented with a display representing data quality as well as strain. The appearance of each image is determined jointly by these two quantities.

The method for spatial averaging is very similar. This is applied to reduce the noise in 3D data spanning a volume. In general, the approach is to convolve the data with a filter kernel, spanning time and the three spatial dimensions. The filter kernel is expressed explicitly in the form $K(|\Delta x|, |\Delta y|, |\Delta z|)$. Spatially averaged data $S(x, y, z)$ and $\Omega(x, y, z)$ are similar to the persisted data: pseudo-strain values after spatial averaging are given by $S(x, y, z)/\Omega(x, y, z)$. These data are calculated as

$$S(x, y, z) = \sum_i K(|x_i - x|, |y_i - y|, |z_i - z|)W_B(i)\epsilon_B(i), \quad (6.7a)$$
$$\Omega(x, y, z) = \sum_i K(|x_i - x|, |y_i - y|, |z_i - z|)W_B(i). \quad (6.7b)$$

This is symmetric smoothing, for which kernel values in the range 0 to 1 express the weighting of normalised pseudo-strain data at $(x_i, y_i, z_i)$ when spatially averaged data are being calculated at $(x, y, z)$. Generally, sets of normalised pseudo-strain $\epsilon_B(i)$ and precision $W_B(i)$ data may be regularly or irregularly distributed over 3D space. Freehand scanning examples in this chapter involve data that are irregularly distributed, but the same interface has been applied in [148] to regularly-sampled data using a mechanically-swept probe.

$K(|x_i - x|, |y_i - y|, |z_i - z|)$ must be defined in a functional form when the location data $(x_i, y_i, z_i)$ are irregular and continuous, rather than as a discrete kernel. Ideally, spatial averaging might be implemented with a smooth kernel, such as a Gaussian, but this is inconvenient to calculate, especially when the data are irregularly distributed. Instead, for later examples a rectangular moving average filter is used,

$$K(|\Delta x|, |\Delta y|, |\Delta z|) = \begin{cases} 
1 & \text{if } |\Delta x| < L \cap |\Delta y| < L \cap |\Delta z| < L, \\
0 & \text{otherwise},
\end{cases} \quad (6.8)$$

where $L$ sets the kernel size. This filter can be applied very efficiently to irregularly or regularly distributed data. The central lobe of its spatial frequency response is a low pass filter. Some high frequency noise remains owing to side lobes, but this can be removed by explicitly low-pass filtering the pixel data before the final display.
6.2.5 Display

Returning to Figure 6.2, an advantage of traditional ultrasound imaging is that signal intensity displays automatically tend to show the most data where the signal is strong, and they show less data where the signal is weak (the image turns black). Similarly, one of the options when imaging pseudo-strain is to control intensity based on the precision data, and to use changes in colour (preferably independently of intensity) to indicate precision data. Options regarding the colour scheme include the use of a wide range of saturated colours, producing the effect of a contour display (e.g., the blue-cyan-green-yellow-red scale of [35] and [40]), but for the present demonstration a dichromatic scale is preferred, which is qualitatively closer to traditional intensity-based displays, and may avoid distorting the features that are perceived to appear in each image.

Aiming for maximum colour variation across the example scale, green and magenta are the extremes, varying from strong green (high strain/soft) through grey (medium strain/medium stiffness) to magenta (low strain/stiff). Humans perceive different colours with different sensitivity, so colour variation at a fixed intensity is achieved following the convention of holding constant the value of

$$59 \times \text{(green pixel value)} + 30 \times \text{(red pixel value)} + 11 \times \text{(blue pixel value)}$$ [189].

The overall colour map, considering both strain and precision, is illustrated in Figure 6.4. The dynamic range for practical examples is set by specifying a top precision level, so strain data are displayed with full intensity when the precision value exceeds the top level, black when precision is near zero, or otherwise with medium intensity.

There are likely to be both advantages and disadvantages associated with representing strain with colour instead of intensity. Image features encoded in these alternative ways are processed with different accuracy and at different speed by the human visual system [190]. Therefore, a second 2D colour map is tested in which strain is indicated by intensity, and a colour (e.g., red) is introduced to indicate precision (see Figure 6.4). Whichever approach is preferable, the aim in relation to displays is to demonstrate that a 2D colour map can be used effectively to depict strain and precision data simultaneously.

Note that in Sections 6.3 and 6.4 the 2D colour maps are encoded with eight bits per pixel. This is usually sufficient to produce good images, because distinctions within the dark or red regions of the colour maps are less perceptible, so these regions can be encoded at low precision. However, 16-bit encoding would be preferable if this type of interface came into widespread clinical use, since the appearance of the display images would then be marginally smoother.

\[^{†}\]High quality colour printing is required in order to fully appreciate the remaining figures.

\[^{‡}\]This is set beforehand to a value that subjectively appears reasonable considering example data sets.
Figure 6.4: Examples of 2D colour maps. (a) Green through to magenta provides the strain scale, while pixel intensity indicates the data quality within a range from a lower threshold — below which everything appears black — up to a maximum threshold — above which colours are displayed with the maximum intensity. (b) The strain scale is based on intensity variation between black and white, which blends with dull red when the precision is low.

6.3 2D results

Consistent analysis parameters and fixed dynamic ranges in the pseudo-strain and precision scales apply throughout the examples in this section. Data for most of the 2D images were acquired using the system described in Section 1.4. Analytic signals produced using FIR filters were converted to baseband envelope/phase representation for WPS, sampled at 13.3 MHz after downsampling. Deformation estimation using the optimised version of WPS from Appendix C was performed with 1D windows of length 1.6 mm (27 samples, 12λ) spaced at intervals of 0.52 mm (9 samples, 4.0λ) down each A-line. A 2D PLLSR kernel of ∼2.0×2.0 mm (4 windows × 7 A-lines) produces strain estimates. The correlation coefficients for predicting precision were calculated in this instance at each final displacement estimate using baseband analytic signals with the same sample rate. The precision values were used to weight the least squares fits of normalisation surfaces and persistence following Equation 6.6 with γ = 0.860.

The scan shown earlier in Figure 6.2 is substantially improved in Figure 6.5 by the full interface scheme with intelligent normalisation, persistence and display. This scan was undertaken using the linear-array probe. Precision-weighted persistence gives rise to good pseudo-strain estimates throughout most of the image. The poor precision of estimates in the shadowed regions is clearly indicated by both of the 2D colour maps.

A Terason\textsuperscript{†} T3000 laptop-based ultrasound machine running Stradwin\textsuperscript{‡} freehand 3D ultrasound software was used to provide an example scanning to greater depth at lower resolution with a convex 5C2 probe of 3.5 MHz nominal centre frequency, from which

\begin{itemize}
  \item[\textsuperscript{†}] http://www.terason.com
  \item[\textsuperscript{‡}] http://mi.eng.cam.ac.uk/~rwp/stradwin
baseband envelope/phase data were sampled at 5.3 MHz after downsampling. The optimised version of WPS was applied with 1D windows of length 2.6 mm (18 samples, 12λ) at intervals of 1.7 mm (12 samples, 8λ) down each A-line. A PLLSR kernel of ∼6.8×6.8 mm (4 windows × 7 A-lines). Persisted images have γ = 0.928.

Figure 6.6 shows images with a maximum depth of 11.9 cm scanning an inhomogeneous gelatin phantom with stiff inclusions. These images are normalised using Equation 6.4. Individual strain images usually produce some regions of good strain estimates, alongside other regions with lower precision. Unweighted frame averaging as in Figure 6.6b might eventually converge on a good image, but for short integration times it is usually less accurate than the best individual images. The advantage of persistence, as in Figure 6.6c–d, is that it makes efficient use of the data, so better strain images are produced easily, with larger regions of good data and generally less noise. In some scans, as in this example, the use of an image-wide weighting is sufficient to cut out most of the noise, though pixel-level weightings often give better results. The other advantage of pixel-level weightings is the retention of pixel-level precision data in persisted images, so it is still possible to indicate data quality using a 2D colour map. However, an image produced using basic tracking (see Chapter 2) is also included in Figure 6.6e, in which the best form of persistence has been employed. Little of the image has sufficient precision to be visible, and where pseudo-strain values are presented they are less accurate. This demonstrates that persistence is far more effective if the rate of severe outliers can be kept to a minimum. Persistence is highly effective in conjunction with algorithms based on robust tracking from Chapter 2 where the implicit continuity constraint avoids bulk errors.

The next example is from freehand scanning of a breast biopsy phantom (Computerised Imaging Reference Systems, Inc.‡ Model 052) using the linear-array probe. The data quality in this case depends less on maintaining even probe pressure, because displacement tracking near to the surface is subject to less motion decorrelation, even if the probe does rotate substantially. However, this means that a wider range of motions register high

\[\text{Figure 6.5: Thyroid images using the full interface.}\]

\[\text{strain 1}\quad \text{strain 2}\quad \text{B-scan}\]

\[\text{†The physical width of this kernel is in fact wider than stated towards the bottom of the image and narrower towards the top.}\]

\[\text{‡http://www.cirsinc.com}\]
Figure 6.6: Examples of alternative persistence methods. (a) **Best individual image without persistence:** Individual frames typically produce a mixture of good and bad image regions with different levels of precision. (b) **Unweighted frame averaging:** The unweighted average of an image sequence is noisier than many of the individual frames. (c) **Precision-weighted frame averaging:** A sequential average weighted by each frame’s mean precision significantly reduces the level of noise. (d) **Pixel-level precision-weighted persistence:** Performing the average with a different weight for every pixel further reduces the level of noise, but only slightly in this example. Its main advantage in this case is the retention of pixel-level persistence data, hence the remaining poor data can be hidden. (e) **Less robust displacement tracking:** This image has the same persistence, but with basic tracking. Persistence is more effective in conjunction with robust tracking.
Figure 6.7: Comparison of normalisation applied to a single frame (no persistence). The normalisation surface is based on (a) Equation 6.2, (b) Equation 6.3 and (c) Equation 6.4.

precision, which actually makes correct normalisation more important than in Figure 6.6. A single frame of strain data with relatively even compression is illustrated in Figure 6.7, exhibiting two noteworthy features. The uniform normalisation in Figure 6.7a gives the impression of there being stiffer material towards the bottom of the image, where the stress disperses into the surrounding material. The images in Figure 6.7b–c are better, because the region with lower stress registers instead as similar pseudo-strain at lower precision, resulting in larger hidden regions. It is also clear from the image that the motion of the probe was slightly rotational, so greater pressure was applied on the right-hand side. This gives an appearance of soft material on the right of the image in Figure 6.7a–b, including a particularly soft region with low precision data. The background material correctly appears more uniform where the more sophisticated normalisation is applied in Figure 6.7c, particularly in the top right of the image, where the data now register an acceptable level of precision.

Rotational movement of the probe often results in stark differences depending on the form of the normalisation. Figure 6.8 shows an extreme example which demonstrates the importance of appropriate normalisation to make best use of the recorded data. The inhomogeneity of pseudo-strain precision in these images means they also highlight the value of correctly applying weighted persistence at the pixel-level. Figure 6.9 shows that precision-weighted frame averaging is no better than unweighted frame averaging in this instance, whereas an excellent pseudo-strain image is produced by applying precision-
Figure 6.8: This image records a frame in which the main motion was rotational, so one side extended while the other compressed. The label 0 denotes the use of a black-white colour map without precision data. (a) If a uniform normalisation is used the resulting pseudo-strain image has one half coloured white and the other black about a pivot. Fortunately the precision data correctly register an absence of useful data, so (a1) and (a2) are blank. However, the more sophisticated normalisation applied in (b) registers many useful measurements, with acceptable precision at the edges of the image, away from a central pivot, the position of which is clearly visible.

weighted persistence at the pixel level. The sophisticated normalisation with lateral stress correction is advantageous, because it reduces the level of noise and produces a pseudo-strain image that corresponds more closely to the stiffness of the phantom material.

Final 2D results in Figure 6.10 show a typical image sequence indicative of the sonographer’s experience when beginning a freehand scan using the new interface. This scan target is an inhomogeneous agar phantom containing half of an olive, which is slightly stiffer than the agar. The screen is initially black (or red) before acceptable data become available. It begins to colour almost immediately on contact with the scan target, although some parts of the image colour less quickly than others, and regions without data such as the shadow on the right remain black/red. Stable images are achieved easily. The development of a successful scanning technique is supported by visual feedback: good technique illuminates the display, whereas poor movements cause it to darken.
Figure 6.9: Examples of persistence alternatives applied to a sequence of strain images from the scan of the breast biopsy phantom, where the scan has been conducted inexpertly, frequently rolling the probe about the elevational axis. (a) Unweighted frame averaging still produces poor results. (b) In this instance, precision-weighted frame averaging is no better than unweighted averaging, because the precision of each individual estimate correlates poorly with the mean precision in each frame. (c) Precision-weighted persistence at the pixel level produces a far better image.
Figure 6.10: Sequence of images at the start of a freehand scan.
Figure 6.11: Freehand scan of an olive-gelatin phantom. Top row: intensity. Middle row: strain. Bottom row: pseudo-strain with spatial averaging \((L=1\text{mm})\). In every row, fine red lines through the middle two slices (axial-elevational and lateral-elevational) indicate their intersections with the axial-lateral slice on the left. Compare previous results based on the same data in Figure 1.17.

### 6.4 3D results

Freehand 3D scans including 6DOF position data were recorded using the system described in Section 1.4 with similar pseudo-strain settings to Section 6.3. Baseband envelope/phase data were downsampled to 11.1 MHz and processed using 1D windows of length 1.8 mm (27 samples, 14\(\lambda\)) spaced at intervals of 0.6 mm (9 samples, 4.9\(\lambda\)) along alternate A-lines, followed by a PLLSR kernel of \(\sim1\times1\) mm (2 windows by 2 alternate A-lines). Where applied, the level of spatial averaging was set to \(L=1\text{mm}\).

Each 3D example requires many images to illustrate the different planes slicing through each volume of data, so to save space only the second 2D colour map (greyscale–red) is shown. Figure 6.11 shows the olive-gelatin phantom scanned during the preliminary investigation of freehand 3D strain imaging [14]. The middle row of Figure 6.11 shows
strain images with a uniform normalisation of each frame, without spatial averaging. The axial-elevational and lateral-elevational planes have been constructed by nearest-neighbour interpolation. This fails to produce a useful 3D image. Previously, the use of frame-level quality measures was investigated for automatic rejection of frames with poor strain images [14, 186]. Now far better results are achieved by applying the new interface with normalisation following Equation 6.4 and pixel-level precision-weighted filtering. One part of the 3D image in Figure 6.11 still produces poor data, because the phantom contained a pocket of trapped air, causing shadowing shown in the axial-elevational slice. The 2D colour map correctly marks the shadowed region red, where no accurate strain data are available.

The spatial averaging incurs minimal loss of resolution, because much of the increase in signal-to-noise ratio arises from the weighted averaging of pseudo-strain data that are spaced more closely than the true resolution of the strain imaging system. This means that weighted spatial averaging is particularly helpful in the context of freehand 3D strain imaging where the data are irregularly distributed, because some strain estimates from successive 2D frames overlap almost precisely in 3D space. This arises because of the relatively haphazard motion of the probe. The errors in overlapping strain estimates may only be weakly correlated, because each overlapping strain estimate arises from a different deformation of the tissue, which makes weighted averaging highly effective. The form of normalisation is also important in this application, because every frame has a different pressure distribution. Effective spatial averaging requires data to have been normalised suitably to adjust for pressure variations.

Another freehand 3D scan in Figure 6.12 shows that the technique is repeatable. This is an olive-agar phantom (the same as in Figure 6.10) which has been scanned more recently than the previous example. Again, spatially averaged pseudo-strain in Figure 6.12 offers a great improvement over naive strain imaging. Readers may also note the benefits of experience with phantom construction: the olive-agar phantom contains half of an olive, upturned to avoid the formation of the air pockets that were problematic in earlier olive-gelatin phantoms [14], hence less of the image is marked red in Figure 6.12.

Can this technique be applied to in vivo scanning? 3D pseudo-strain imaging has not yet been performed on patients, but the clinician working in Cambridge on clinical applications for pseudo-strain imaging has on one occasion attended the laboratory to attempt freehand 3D scanning of a healthy volunteer’s thyroid. Figure 6.13 shows the result. The image provides a 3D display that appears plausibly related to likely mechanical properties of the scan target, with slip planes above and below the thyroid, which itself is somewhat softer than surrounding structures. Clinical experience will be required to find applications where 3D pseudo-strain imaging is actually useful.
Figure 6.12: Frehand scan of the olive-agar phantom. Top row: intensity. Middle row: strain. Bottom row: pseudo-strain. In every row, fine red lines through the middle two slices (axial-elevational and lateral-elevational) indicate their intersections with the axial-lateral slice on the left.

6.5 Conclusion

A novel interface has been presented for real-time 2D and tracked 3D freehand pseudo-strain imaging, with brief explanations of the underlying theoretical principles. Preferred inputs for the interface are strain estimates from a robust strain estimator with accurate precision data. The interface can be incorporated as the front end on a wide range of strain imaging systems, but the best results seem likely to be achieved in systems that neither rely on exhaustive searching nor on tracking methods that exhibit excessive fragility.

Notable aspects of the interface include a normalisation stage, persistence or spatial averaging, and novel display with a 2D colour map. Normalisation reduces the ambiguity of strain data, and actually reduces the level of noise in persisted images. It follows that good, informative pseudo-strain images can be produced by a wide range of probe motions, rather than relying heavily on careful, even compression. In order to exploit
these benefits fully, persistence needs to be weighted at the level of individual pixels, rather than at the level of sequential images.

The interface can just as well be applied to freehand 3D scanning as to real-time 2D scanning. The acquisition of 3D volumes of pseudo-strain data is essentially no more challenging than the basic problem of reliably producing good 2D images, and a 2D display no longer flickers depending on the instantaneous motion. Successful 3D scanning is also made easier by the presentation of persisted real-time pseudo-strain images in 2D while the sonographer sweeps out a volume. In general, this system not only improves the quality of the results from particular data sets, but it can also help the sonographer to develop a successful scanning technique.

It is already easy to produce fairly good pseudo-strain images for simple targets using the current version of the interface. There are no doubt many ways of improving on the performance by modifying implementational details while continuing with the same general principles. For example, there may be scope for testing how the absolute performance is affected by different types of precision estimates. Limitations of the precision estimates used in this chapter were discussed. Alternatives to correlation-based precision estimators may perform better [171]. There is also scope for devising different normalisation surfaces, which might be target-specific. If a large number of parameters are required, it is worth noting that complicated normalisations can be produced at lower computational cost by fitting surfaces to strain rather than displacement values [148].
Part IV
Chapter 7

Conclusion

Several approaches to ultrasonic elasticity imaging were discussed in Chapter 1. The potential advantages of quasistatic methods include high resolution for depicting the geometry of tissue regions distinguished by their relative mechanical properties. Widespread clinical benefits are likely if a practical form of quasistatic elasticity imaging is developed and implemented on general-purpose ultrasound machines. The freehand scanning technique must produce meaningful images without being too difficult or labour-intensive for sonographers. It will be all the more valuable if the technique is simple to implement across a wide range of hardware platforms.

Various new methods have been presented in this thesis which may support useful elasticity imaging with an easy scanning technique. Pseudo-strain extends strain imaging by recognising the need to consider the level of applied stress in order for strain data to be meaningful, even if pseudo-strain images do not correspond directly to any particular material property. The pseudo-strain approach facilitates more effective post-processing to produce more informative images.

Part II presented efficient methods for producing robust deformation estimates that are highly accurate in the context of past literature, surpassing the performance of relatively elaborate alternatives under normal scanning conditions. Optimisation was also discussed, since fast processing is important for producing a live display during freehand scanning. Part III focused on strategies for exploiting deformation estimates in combination with precision indicators. The analysis parameters can be modulated to achieve fairly uniform precision at nonuniform resolution. Methods similar to DRS may yet be worth incorporating in practical imaging systems, but the fusion of data from multiple images seems to be of more immediate importance. Variable normalisation followed by precision-weighted persistence or filtering provides superior pseudo-strain and precision data.

A broad-focus clinical study funded by the Wellcome Trust† began mid-2007 at Addenbrooke’s Hospital (Cambridge, UK) testing pseudo-strain imaging based on the interface from Chapter 6 with the displacement estimator from Chapter 4, location estimation from Chapter 3, and tracking similar to the strategies in Chapter 2. Pseudo-strain imaging is

†Translation Award 081511/Z/06/Z.
being performed alongside miscellaneous routine ultrasound scans to identify applications where it may help. Scans with position sensors for 3D pseudo-strain imaging have not yet been undertaken in a clinical setting, but roughly 80 patients have so far been scanned in 2D examinations, mostly of the head and neck. The early images are encouraging. Later stages of the study will focus on the targets that seem most promising.

Example images in Figures 7.1 and 7.2 are taken from scans performed using a Tera-son T3000 laptop-based ultrasound machine with a 12L5V probe of centre frequency 7.75 MHz sampled at 40 MHz, from which baseband phase/envelope data were produced as described in Chapter 4 and downsampled to 13.3 MHz, before WPS deformation estimation with AMC using 1D windows of length 1.6 mm (28 samples, 18 λ) spaced at intervals of 0.5 mm (9 samples, 6 λ), followed by PLLSR strain estimation using a 2D kernel of ∼2.0×2.0 mm (4 windows × 7 A-lines). The 2D pseudo-strain colour map has the same precision scale as examples in Chapter 6 and the persistence setting is γ = 0.860.

The pseudo-strain image in Figure 7.1b shows a ∼2 cm thyroid nodule (follicular adenoma) in a 53 year old woman. A dark patch indicates a soft interior surrounded by stiff (white) material. This correlates with the typical structure of a follicular adenoma, consisting of soft tissue inside a thick fibrous capsule [191].

It is interesting to note changes when the imaging analysis is altered. Following Chapter 4, little difference is expected between images produced by current forms of WPS and EPZS with AMC. However, Chapter 3 showed that omitting AMC can have a stark impact on image quality if strain images are produced by differencing displacement estimates from long windows. Images in Figure 7.1 were produced with substantial post-filtering (a large 2D PLLSR kernel and persistence) which reduces the significance of AMC, but there are still subtle changes when AMC is omitted in Figure 7.1c. AMC is not essential for producing useful pseudo-strain images, but images with AMC such as Figure 7.1b are somewhat more accurate, and hence potentially more valuable.

Figure 7.1d–g shows that robust tracking and a good interface are essential to the feasibility of producing useful pseudo-strain images of this thyroid nodule. The change in Figure 7.1d is simpler normalisation following Equation 6.2. Normalisation modifies both the values and the precisions of pseudo-strain data, so clear differences between the images are unsurprising. The simpler normalisation is expected to represent less accurately what can reasonably be inferred from freehand deformations, resulting in a higher level of noise. In this instance both Figure 7.1b and Figure 7.1d have plausible appearances; the “ideal” image is unknown, so the relative performance of the normalisation strategies cannot be inferred. The use of basic tracking instead of robust strategies incurs a more obvious reduction in performance in Figure 7.1f.

The most striking differences arise when the persistence method is changed. Figure 7.1e demonstrates unweighted persistence. Vertical streaks in Figure 7.1e1 arise because the more sophisticated normalisation cannot sensibly be used without variable per-

---

1Ethical approval for these scans was granted by Cambridgeshire 3 Research Ethics Committee: “A pilot study to evaluate the clinical application of ultrasound elastography”, ref. 07/H0306/90.

2http://www.terason.com
Figure 7.1: Large thyroid nodule (follicular adenoma) in a 53 year old woman. (a) B-scan. (b) Standard image by method described in text. Other images omit aspects of the full pseudo-strain analysis. (c) AMC not applied. (d) Single normalisation value per frame. (e) Unweighted persistence with (e1) sophisticated normalisation and (e2) single normalisation value. (f) Basic tracking. (g) Persistence weighted on the per-frame level with (g1) full normalisation and (g2) single normalisation value.
sistence, since the pseudo-strains have extremely low precision where the normalisation strain approaches zero. The streaks do not appear in Figure 7.1e2, where unweighted persistence is combined with simple normalisation. Nonetheless, this image is equally uninformative. Weighting each frame according to the frame-wide mean precision achieves somewhat better results. Sophisticated normalisation in Figure 7.1g1 is still inapplicable, but the result with simple normalisation in Figure 7.1g2 is partially successful. Nevertheless, it is far less successful than pixel-level persistence, and may not be superior to inspecting a sequence of pseudo-strain images without persistence to select a relatively successful individual image. The loss of pixel-level precision data is also a severe disadvantage, because the masking out of low-precision data helps to make the images more informative.

The other example in Figure 7.2 shows the parotid gland of a 22 year old woman. A stiff ~5 mm tumour in the top-centre of the image was found on histology to be a benign pleomorphic adenoma. It is barely visible in the B-scan as a slightly hypoechoic patch with posterior echo enhancement, whereas it appears clearly in the pseudo-strain image as a stiff lump. Two patches appear stiff, one of which is caused by a stationary reverberation pattern, so it is useful to examine the B-scan for reference. Interestingly, this example depends less on the new analysis methods. The scattering field is fairly uniform, without notable decorrelating features, so there is relatively little change depending on the use of AMC, sophisticated normalisation and robust tracking, but the persistence method is again important. Unweighted persistence is a failure, but frame-level weighting in Figure 7.2g2 produces a good image in this case. The level of precision would seem here to vary in a fairly uniform pattern between one 2D frame and the next, so pixel-level weighting is not essential, although pixel-level precision data are still useful for masking inaccurate pseudo-strain values.

To summarise, pseudo-strain imaging is a practical technique that can provide valuable additional information relating to mechanical properties alongside conventional ultrasound imaging. Persistence in 2D displays and spatial filtering in 3D scans seem particularly important for producing stable, intelligible images, so the interface from Chapter 6 may significantly increase the feasibility of clinical pseudo-strain imaging. The two examples in this chapter provide a limited indication of the true value of each method. This does not supersede the findings of analysis and experiments in the earlier chapters, although it is interesting to note the different sensitivities in each example to changes in the imaging analysis.

Different aspects of the analysis might be more important in other clinical examples. Following Chapter 3, AMC is expected never to reduce the accuracy of displacement estimates, and sometimes it may yield substantial improvements. However, AMC only makes a significant difference if displacement estimation is performed using windows of substantial size compared to the level of blurring associated with subsequent filtering. Given that the distinction is potentially subtle, it is convenient that AMC is far less computationally costly than the previous alternative of adaptive stretching. The next contribution of this thesis was a general discussion of phase-based estimators including WPS, which usefuly
advances an understanding of how these estimators work, providing a novel framework vis-à-vis correlation analysis which may stimulate further development. Similarly, the work on dynamic resolution selection served primarily as groundwork for understanding the statistical properties of strain estimates, supporting the interface concept. The interface probably is the single most important contribution of this thesis, combining sensible normalisation schemes with filtering weighted at the pixel level, which makes it far easier to perform successful pseudo-strain imaging. Additionally, robust tracking is important in this context, because the images need to have reliable properties with few bulk errors to harness the full benefits of weighted filtering.
CHAPTER 7. CONCLUSION

Future work

All aspects of ultrasound scanning and pseudo-strain imaging depend partly on the scan target. The suitability of specific modifications to the tracking, filtering and normalisation schemes is ultimately unknown prior to clinical testing, so feedback between the clinical and technical strands of future work should be beneficial. Some features of the imaging analysis may be advanced for general purposes, while other aspects such as normalisation may be improved by task-specific modifications relating to the properties of specific tissues.

Pseudo-strain imaging as described in this thesis has some attractive properties. It seems certain that some soft tissues can be understood better in medical examinations by applying the technique in its current form, but no doubt there remains scope for improvement. Regardless of which applications may be found in the near term, the range of applications and the general utility of pseudo-strain imaging may be improved by further technical development.

A few suggestions are now made for strands of future development that, in the author’s opinion, are likely to be worthwhile. These strands may not be intuitively obvious. The themes outlined in Section 1.5 will remain relevant, the more important of which are maintaining “known accuracy” followed by “creative exploitation” of the combined deformation and precision data. Developments can be made both in deformation estimation and in the interface (or post-processing).

Chapter 1 mentioned many applications for deformation estimation aside from pseudo-strain imaging. However, deformation estimation may not be the most important area for development in the pseudo-strain imaging context. In the past, accurate estimation of displacement and strain using large windows of RF ultrasound data has been a popular research topic. Location estimation in AMC continued this line of enquiry. However, performance has commonly been assessed in terms of the precision achieved when imaging uniform strain fields, i.e., $\text{SNR}_{e}$, which may bias the choice of analysis methods towards large windows, while paying less attention than appropriate to the reduced resolution shown in Chapter 5 and the distortion of features where strain is inhomogeneous [171]. For further development it may be interesting to use much smaller windows combined with optimised filtering methods [170, 171].

Why have windows of substantial size been used so often throughout this thesis and in past studies? One reason is that exhaustive search depends on having enough data in each window to avoid unacceptable rates of bulk errors, and it is also easier to devise robust tracking methods when the windows are fairly large. However, by further improving robust tracking it may be possible to avoid bulk errors using much smaller windows, perhaps down to the level of individual RF ultrasound samples. The second reason for using large windows is that, provided AM does not introduce noise, the strain estimates are more accurate prior to filtering, as predicted by Appendix A.1 and demonstrated throughout Chapters 3 and 4. However, it may be that the estimates from small windows

---

†This comment is based on personal communications with Dr. Susan Freeman (radiologist).
lead to more accurate strain images that better represent the structure in pseudo-strain fields if the subsequent filtering methods are improved [170, 171]. Further research is warranted considering robust tracking and filtering methods with a view to using very small windows.

Intuitively, improved estimation of displacement and strain in the lateral direction seems another appropriate goal for future work. However, while lateral strain estimates are certainly useful in related applications such as poroelasticity imaging, it is by no means self-evident that lateral strain estimates would provide valuable additional information for pseudo-strain examinations of the many tissues that are near-incompressible. From a technical perspective, it seems likely that lateral strain estimation may require substantial modification to beamforming methods to achieve a useful level of accuracy [28, 192]. The primary value of lateral tracking may be consequent improvement in the accuracy of axial strain estimates. In that case, the aim of further research would be to improve axial strain estimates beyond what is already achieved by coarse lateral tracking. It may be that other research topics lead more directly to improvements in pseudo-strain imaging.

Work on details of the interface is likely to remain important in the immediate future to further enhance clinical utility. The method for assessing precision in Chapters 5 and 6 can be improved. For example, the present method is fairly accurate assuming that bulk errors never occur, but nonlinear techniques to adjust precision data based on the likelihood of bulk error would be useful. The varied properties of clinical ultrasound data make it unlikely that bulk errors can be eliminated throughout every image. Additionally, the use of very small windows and optimised filtering in future imaging analysis might change the suitability of different methods for assessing precision.

Normalisation is another important component in the extraction of meaningful pseudo-strain data from deformation estimates, especially when used in conjunction with persistence or spatial filtering. Pseudo-strain imaging was motivated in this thesis as a compromise between basic strain imaging and tackling the inverse problem. No attempt has been made at solving the inverse problem, but theory motivated by inverse problems may be valuable from the perspective of pseudo-strain imaging as a means of devising superior normalisation strategies. In particular, it should be possible to adjust the form of the normalisation surface based on knowledge of the typical supporting anatomy surrounding specific tissue targets.
Appendix A

Derivations relating to AMC

A.1 Relationship between SNR_e and window length

This derivation concerns the likely trend in strain estimation performance as a function of window length when AM is negligible or AMC is perfect. When location errors are zero, Equation 3.5 reduces to

$$\text{SNR}_e = \sqrt{\frac{\varepsilon^2 \Delta y^2}{\sigma_D^2}}. \quad (A.1a)$$

Following the definition of $\hat{D}$ in Equation 3.2a it follows that

$$\sigma_D^2 = \sigma_{\hat{u}_1}^2 + \sigma_{\hat{u}_2}^2 - 2\sigma_{\hat{u}_1\hat{u}_2}. \quad (A.1b)$$

The reciprocal of $\sigma_{\hat{u}_1}^2$ is the displacement estimation precision, which is proportional to the information content of the window. In the absence of severe data misalignment (bulk and phase-wrapping errors) the information content is assumed directly proportional to the window length, with a constant of proportionality, $K_i$.

$$\sigma_{\hat{u}_1}^2 = \sigma_{\hat{u}_2}^2 = \sigma_{\hat{u}}^2 = \frac{1}{K_i Y} \quad (A.1c)$$

Next, the correlation between successive errors is assumed equal to the displacement estimation variance multiplied by the fraction of data that is common to both windows:

$$\sigma_{\hat{u}_1\hat{u}_2} = \frac{1}{K_i Y} \times \left( \frac{Y - \Delta y}{Y} \right). \quad (A.1d)$$

Combining (A.1a)–(A.1d) yields a proportional relationship between SNR_e and $Y$:

$$\text{SNR}_e = \left( \varepsilon \sqrt{\frac{1}{2} K_i \Delta y} \right) Y. \quad (A.2)$$

A straight line on a graph of SNR_e against window length therefore indicates that location errors are very small. Location error is not the only possible cause of perturbation away from a straight line; long windows may also be affected by phase-wrapping at window edges, in which case Equation A.1c becomes inaccurate. Furthermore, Equation A.2 does not imply that SNR_e is proportional to strain, because a change in $\varepsilon$ affects the level of decorrelation, thereby altering $K_i$. 
A.2 AMC for a correlation coefficient variant

The real part of the complex correlation coefficient, \( R_{a_1a_2} \), is

\[
R_{a_1a_2}(n\Delta y, \tilde{u}) = \Re \left[ \frac{\sum_{n\Delta y}^{n\Delta y+Y} a_1^{*}(y)a_2(y+\tilde{u})}{\sqrt{\sum_{n\Delta y}^{n\Delta y+Y} |a_1(y)|^2 \sum_{n\Delta y}^{n\Delta y+Y} |a_2(y+\tilde{u})|^2}} \right]. \tag{A.3}
\]

Symbols defined in Section 3.2 have the same meanings here. \( R_{a_1a_2} \) can be used as an alternative correlation coefficient, so displacement estimates are identified at peaks:

\[
\hat{u}_n = \arg \max_{\tilde{u}} R_{a_1a_2}(n\Delta y, \tilde{u}). \tag{A.4}
\]

This expression will be analysed following the method described in Section 3.2.2, considering only the common component from the signal model in Equation 3.12. At an estimate, \( \hat{u}_n \), the derivative of \( R_{a_1a_2} \) with respect to \( \tilde{u} \) must be zero. A considerably simpler approximation can be made by retaining terms that include the derivative of the underlying periodic signal, but neglecting the derivative of its envelope:

\[
\frac{d}{d\tilde{u}} R_{a_1a_2}(n\Delta t, \hat{u}_n) = 0 \simeq \Re \left[ \frac{\sum_{n\Delta y}^{n\Delta y+Y} f(y)e^{-j\phi(y)} f(y_2)e^{j\phi(y_2)} j\frac{df(y_2)}{dy_2}}{\sqrt{\sum_{n\Delta y}^{n\Delta y+Y} f(y)^2 \sum_{n\Delta y}^{n\Delta y+Y} f(y_2)^2}} \right]. \tag{A.5}
\]

The numerator in this expression must itself be equal to zero. The derivative effectively multiplies each term in the summation by \( j \) times the local frequency. Neglecting local frequency changes (following an argument put in Section 3.2.3) the expression can be simplified further:

\[
\Re \left[ \sum_{n\Delta y}^{n\Delta y+Y} f(y)f(y_2)e^{j(\phi(y_2)-\phi(y)+\frac{\pi}{2})} \right] \simeq 0,
\]

\[
\therefore \sum_{n\Delta y}^{n\Delta y+Y} f(y)f(y_2) \sin(\phi(y_2) - \phi(y)) \simeq 0. \tag{A.6}
\]

This is the same as Equation 3.16 in Section 3.2.3, from which the rest of the argument follows identically, leading to weightings based on the product of pre- and post-deformation envelopes. This is a second example of the analytical method for deriving weighting approximations for AMC. Nevertheless, the fact that a heuristic weighting approximation for the traditional correlation coefficient is tested in Chapter 3 does not necessarily imply that performance would be superior using the real part of the complex correlation coefficient.
Appendix B

Adaptive strain estimation

Previous adaptive strain estimators (ASEs) have used correlation coefficient [136] or SAD [137] as the measure of signal similarity. Both options are now tested. Typically a window of pre-deformation data is matched with a post-deformation window that is shifted and stretched, requiring searches over both displacement and stretch [136, 175]. Previous papers have presented minimal details regarding the two-dimensional search, although details of the algorithm can have a substantial impact on the overall estimation accuracy.

B.1 Search methods

Exhaustive search methods, while generally robust, are prone to extremely large outlier errors. This is particularly problematic when searching exhaustively over window displacement, which is enabled in correlation coefficient methods by applying the Fast Fourier Transform for relatively efficient calculation [136]. The wider the exhaustive search, the greater the risk that a distant outlier has a high correlation, by chance, resulting in bulk displacement error. These errors are present, for example, in several figures presented for a qualitative assessment in [136]. It is important to eliminate these bulk errors in following quantitative tests, however, because they severely reduce SNR, so that the general algorithm performance is poorly represented. Therefore, general robustness is achieved here by a multi-stage semi-exhaustive search.

Stage 1 searches only over stretch. The mechanics are illustrated in Figure B.1 where \( u_{os,n} \) denotes the offset displacement at the beginning of window \( n \), and \( \tilde{\varepsilon}_n \) denotes a trial stretch factor. \( \tilde{\varepsilon}_{n-1} \) denotes the final stretch of the preceding window, from which a fixed point is used as a pivot, about which both the offset displacement of window \( n \) and the relative displacements of its contents are stretched until an optimum has been found (maximum correlation coefficient or minimum SAD). Preliminary tests on uniform strain data indicated that a pivot at the start of the preceding window (as illustrated) yields best results. The search is conducted by interval reduction, starting with upper and lower trial stretches of \( \pm 10\% \), and evaluating the objective function (correlation coefficient or SAD) for both cases. The poorer trial stretch is replaced by a new value from within the search interval, so the interval width shrinks by 10\% at every iteration. Stage 1 terminates when the interval is smaller than 0.01\%. 

161
Figure B.1: Stage 1 in ASE. The offset displacement at the start of window \( n - 1 \) serves as a pivot for a search over stretch at window \( n \). Trial stretches \( \varepsilon_n \) change both the offset displacement at the start of the window and the relative displacements of its contents.

Although Stage 1 is not a fast searching method, it avoids the higher order computational complexity of alternative two-parameter search methods, by limiting the search space to a single parameter. Stage 1 can be applied in isolation, but the disadvantage then is that estimates from consecutive windows are highly inter-dependent. The stretch factors therefore have considerably lower resolution than in methods where windows are treated independently [136]. Differencing of window centre displacements following Stage 1 nonetheless offers more comparable resolution, but for a meaningful comparison with other algorithms it is preferable to use Stage 1 only for global robustness, in combination with a subsequent stage that offers greater independence in the fine scale estimates of consecutive windows.

Stage 2 refines the output of Stage 1 with a local search over displacement and stretch. Stage 2a does not change the stretch factor, but it applies an uphill search to locate integer sample offset displacements to either side of the optimum, starting from the offset displacement of the previous stage. Then binary interval reduction is applied to find the subsample optimal offset displacement. Stage 2b then searches over the stretch factor, taking the centre of the current window as the pivot. This stretching search is similar to the interval reduction method of Stage 1, but it begins with a stretching interval \( \pm 2\% \) to either side of the previous estimate. The mechanics of Stage 2 are illustrated in Figure B.2. Stages 2a and 2b loop iteratively until a convergence criterion is satisfied.

It will be noted that every calculation in the search requires many instances of signal interpolation. This is performed by the high accuracy method expressed in Equation 3.32 for a fair comparison with the other algorithms. Additionally, results can be distorted if the search guidance in Stage 1 fails, which can happen if there are small regions of high decorrelation. However, error propagation is avoided by initialising Stage 1 with the offset displacement from a neighbouring vector if the neighbour had a better objective function at the preceding window, i.e., cross-seeding as described in Chapter 2.
Figure B.2: Illustration of a single iteration of Stage 2 in ASE. This comprises an uphill search for the optimal offset displacement in Stage 2a, followed by interval reduction to identify the optimal stretch factor in Stage 2b.

B.2 Strain calculation

Two different approaches are possible for the strain calculation. Firstly, displacements of the window centres can be used for gradient-based strain estimation, as with the other algorithms. This represents a fair comparison, because all of the algorithms are tested at the same strain processing resolution, dictated by the window spacing.

Strain estimation by differencing closely-spaced window displacements is used in the tests, because it enables meaningful performance comparisons, in which the strain estimates are sensitive to displacement estimation noise. Differencing, as in Equation 3.1, scales displacement estimation error by the reciprocal of the window spacing. The use of longer windows reduces the maximum achievable resolution, so for practical purposes it makes more sense to difference long windows over a greater separation, or to apply more sophisticated gradient estimation filters such as least squares [163] or wavelet analysis [193]. This reduces strain estimation noise, and it incurs no additional reduction in the overall resolution, provided that the length of the strain estimation filter does not exceed a critical fraction of the window length (see Chapter 5). However, the comparative tests in Chapter 3 are performed deliberately with a fixed strain processing resolution, to obtain results that unambiguously relate to the accuracy of the estimated displacement field.

When using ASE, the alternative approach to strain estimation is the adoption as strain estimates of the stretch factors that were applied to the post-deformation windows. This latter option is elegant [136], but its strain processing resolution is dictated by the window length, so it does not represent a fair benchmark. The performance appears disappointing when using short windows (because the strain processing resolution is effectively lower than in the gradient-based techniques) and encouraging when using long windows (because both the displacement and strain processing resolutions have been changed, whereas only displacement processing resolution changes in the other
algorithms). However, it is unlikely that either of these cases indicates any substantive performance difference as against gradient methods when considering suitably optimised practical systems.

## B.3 Evaluation

The experiments of Figures 3.5, 3.7 and 3.14a are repeated in Figure B.3 with a set of four ASE variants. For brevity, correlation coefficient searches are denoted by the suffix C, and SAD searches by S. Suffix 1 indicates that only Stage 1 is applied, whereas suffix 2 indicates the full two-stage search. The complete list of combinations is ASE_C1, ASE_S1, ASE_C2 (providing benchmark data in the main text) and ASE_S2.

The results show that correlation coefficient outperforms SAD, although it is less easy to differentiate the relative merit of Stage 1 in isolation versus the two-stage search. The two-stage search of ASE_C2 is taken as the benchmark in Chapter 3, because ASE_C1 has a relatively coarse correction, as illustrated in Figure B.4, so it is likely to offer poorer SNR compared to estimators with AMC or amplitude compression when resolving changes in strain at a medium or small scale.

However, it is still interesting that Stage 1 in isolation offers competitive performance, since it indicates an appropriate approach to implementing ASE for practical scanning purposes. Although Stage 1 is far from being a real-time algorithm, the reduction to a single search dimension means that ASE_C1 requires orders of magnitude fewer computations than ASE_C2. Approaches similar to ASE_C1 have been pursued by other authors for algorithm optimisation [194, 195].
Figure B.4: Correction resolution in variants of ASE. These strain images show stretch factors rather than the gradient of displacement estimates. Simulation data at 20 dB were generated with 1% compression in the background and zero strain in inclusion bands of width 0.625 mm. (a) shows the ideal strain image, (b) shows the stretch factors of ASE_C1 windows using $Y = 15\lambda$ and (c) shows the stretch factors of ASE_C2 windows of the same length. The stretch factors of ASE_C1 windows are inter-dependent, which lowers the resolution, so although the correction may be good in homogeneous regions, AM noise will be handled less well at inhomogeneities. By contrast, the stretch factors of ASE_C2 windows are slightly less accurate in homogeneous regions, but they respond better to inhomogeneities, since they depend only on within-window data.
Appendix C

Computational load of phase-based estimators

This appendix discusses computational considerations and simple optimisation techniques. The computational load arises from pre-processing of the ultrasound signals as well as deformation estimation itself. Contributions to the load are listed in Table C.1, quoting values for example implementations of WPS and EPZS. Operation counts are expressed in terms of the number of analytic signal samples in each ultrasound frame, $N$, the length of an FIR filter used to calculate real and imaginary analytic datapairs, $F$, and the number of samples in each deformation estimation window, $M$. The total computational load per frame is the per-frame cost of pre-processing, plus the number of windows multiplied by the per-window cost of deformation estimation. Assuming that ultrasound frames arrive as a sequence, each new frame of deformation data only requires a single new frame of RF ultrasound to be pre-processed.

Performance results for Section 4.4 were recorded using naive implementations with the maximum sample rate and 64-bit floating point arithmetic, compared to which some aspects of the following optimisation measurably alter performance, as will be mentioned. The WPS example implements any form of WPS with amplitude weightings, phase weightings and AMC (the cost is actually lower if any of these is omitted, such as in WPS_L0). The EPZS example implements EPZS_A1. The cost of EPZS_A1 is lower than discarded amplitude variants, which require extra pre-processing to flatten the envelope. The cost can be reduced further by omitting AMC, but the performance is then substantially worse.

C.1 Pre-processing

With the hardware of Section 1.4, real-time strain imaging is achieved by sampling the raw RF signal at 66.7 MHz. This is converted into an analytic signal, optionally at a lower sampling rate, by applying a pair of matched bandpass filters with linear phase, one of EPZS_A2 would be a less helpful example, because it requires phase data as well as real and imaginary data, so it is most efficiently “implemented” by emulation through the WPS framework. Anyway, EPZS_A2 is only a small improvement beyond EPZS_A1, which was tested extensively in Chapter 3.
Table C.1: Contributions to computational load.

$N =$ number of analytic signal samples per frame  
$F =$ FIR filter length  
$M =$ number of samples in each window

<table>
<thead>
<tr>
<th>Task</th>
<th>Computational load</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-processing</strong> (per frame)</td>
<td></td>
</tr>
<tr>
<td>Hilbert filter (both WPS and EPZS)</td>
<td>$2FN$ multiplications, $2(F-1)N$ additions</td>
</tr>
<tr>
<td>Acquire envelope (WPS)</td>
<td>$2N$ multiplications, $N$ additions, $N$ square root</td>
</tr>
<tr>
<td>Acquire phase (WPS)</td>
<td>$N$ atan2</td>
</tr>
<tr>
<td>Convert phase to baseband unsigned integer representation (WPS)</td>
<td>$N$ multiplications, $N$ additions</td>
</tr>
<tr>
<td>Baseband conversion (EPZS)</td>
<td>$4N$ multiplications, $2N$ additions</td>
</tr>
<tr>
<td>Accumulate squared envelope (EPZS)</td>
<td>$2N$ multiplications, $2N$ additions</td>
</tr>
<tr>
<td>Acquire envelope (EPZS with AMC)</td>
<td>$N$ square root</td>
</tr>
<tr>
<td>Each Type A iteration (WPS)</td>
<td>$7M + 4$ multiplications, $12M + 3$ additions, $2M$ bit-compare, $6M$ table-lookup, 1 division</td>
</tr>
<tr>
<td>Each Type B iteration (WPS)</td>
<td>3 multiplications, 2 additions, 1 division</td>
</tr>
<tr>
<td>Extra calculation for location estimate (WPS)</td>
<td>1 division</td>
</tr>
<tr>
<td>Extra calculation for accuracy indicator (WPS)</td>
<td>2 multiplications, 1 addition, 1 division</td>
</tr>
<tr>
<td>Each Type A iteration (EPZS)</td>
<td>$8M + 5$ multiplications, $8M + 2$ additions, 1 atan2, 1 phase-unwrap</td>
</tr>
<tr>
<td>Each Type B iteration (EPZS)</td>
<td>5 multiplications, 2 additions, 1 atan2, 1 phase-unwrap</td>
</tr>
<tr>
<td>Extra calculation for location estimate (AMC for EPZS)</td>
<td>$4M + 4$ multiplications, $4M + 2$ additions, 1 division</td>
</tr>
<tr>
<td>Extra calculation for accuracy indicator, i.e., correlation coefficient (EPZS)</td>
<td>8 multiplications, 9 additions, 1 division</td>
</tr>
<tr>
<td><strong>Deformation estimation (per window)</strong></td>
<td></td>
</tr>
<tr>
<td>Cross-seeding (see Chapter 2): initialising each search with a best displacement estimate from the previous row. Occurs once per window.</td>
<td></td>
</tr>
<tr>
<td>Continuity checking (see Chapter 2): ensuring that the better estimate from two passes has been selected. Only required after multi-pass analysis. Occurs once per window.</td>
<td></td>
</tr>
</tbody>
</table>

which is symmetric (for the real part), the other being anti-symmetric (for the imaginary part). These are implemented in software as 31-point FIR filters ($F = 31$), producing each analytic sample at a cost of 62 multiplications and 60 additions.

This and all subsequent costs scale with the analytic sampling rate. Dense sampling
of the raw RF suppresses electrical noise, which would otherwise be aliased. During filtering there is scope for down-sampling, incurring no loss of information, provided that the new sampling rate exceeds twice the full bandwidth of the filtered signal. In fact, the algorithms described in Part II deteriorate measurably before the sampling limit, because linear interpolation of envelope and baseband phase (WPS) or baseband analytic signals (EPZS) gradually becomes less accurate, although in this system down-sampling to 13.3 MHz incurs less than a 5% reduction in SNR, while conferring an almost fivefold reduction in the overall computational load.

For WPS, a further pre-processing cost is the conversion from analytic signals to envelope and phase data. The implementation here employs the most obvious method: real and imaginary parts are squared, summed and square-rooted for the envelope; a standard atan2 implementation produces phase, which is converted to baseband by subtracting phase offsets. For efficient processing (explained below) this is scaled and converted to unsigned integer representation — low-order bits encode $-\pi$ to $+\pi$, while high-order bits and overflow correspond to $2^n\pi$ offsets. This increases quantisation noise, but 12-bit encoding of $-\pi$ to $+\pi$ has been found to incur no measurable reduction in performance.

EPZS has different additional costs. A substantial number of multiplications are required to demodulate the analytic signal to baseband. Additionally, squared envelope values are accumulated over the image to enable quick normalisation of correlation coefficients. The envelope itself is required as well for location estimation.

C.2 Deformation estimation

At each window the computational load of deformation estimation arises mostly from iterative displacement estimation, although the costs of one-off calculations for a location estimate (for AMC) and an accuracy indicator (for robust tracking) are not negligible.

For both WPS and EPZS, the use of linear interpolation enables rearrangement of the key expressions to achieve extremely efficient iterative processing. Significantly, most iterations on the lines of Equation 4.13 in fact refine subsample displacement estimates that have already converged in terms of integer shifts.

Consider an EPZS iteration. Linear interpolation of the post-deformation baseband analytic signal is used to calculate the complex cross-correlation function at a subsample shift. The trial displacement is an integer shift, $U_k$, plus a fraction, $\delta_k$, i.e., $\tilde{u}_k = U_k + \delta_k$.  

---

1 The full bandwidth may exceed the nominal bandwidth. Sampling at barely twice the -3dB bandwidth incurs considerable aliasing if a small but substantial fraction of signal power is associated with frequencies outside the band that is sufficiently sampled.

2 A lower sampling rate could be used with precise band-limited interpolation, but this might increase the overall computational load, owing to the greater complexity of the interpolation method.
so the practical calculation (cf., Equation 3.10a) is

\[
\langle a_1, a_2 \rangle (n\Delta y, \tilde{u}_k) = \sum_{y=n\Delta y}^{n\Delta y+Y} a_1^*(y) \left( (1 - \delta_y)a_2(y + U_k) + \delta_y a_2(y + U_k + 1) \right) 
\]

(C.1a)

\[
= (1 - \delta_y) \sum_{y=n\Delta y}^{n\Delta y+Y} a_1^*(y)a_2(y + U_k) + \delta_y \sum_{y=n\Delta y}^{n\Delta y+Y} a_1^*(y)a_2(y + U_k + 1) 
\]

(C.1b)

This is a useful rearrangement, placing interpolation factors outside the summations. On every first iteration, and whenever \( \tilde{u}_k \) moves between a new pair of integers, new summations are evaluated and stored ("Type A iterations"). The operation count is reduced dramatically by reusing stored summations whenever \( \tilde{u}_k \) lies between the same pair of integers as the preceding iteration ("Type B iterations"). Most iterations are Type B. The average number of Type A iterations per window is usually barely greater than 1, although it can increase to 2 at high strains with a high sampling rate.

WPS also uses linear interpolation, but it is complicated by the need for correct phase-wrapping. The rigorous method applied in Chapter 4 consisted of wrapping consecutive phase values to each other, then interpolating linearly, and subtracting pre-deformation phase, before finally re-wrapping to the range specified in Section 4.3.1. Calculating phase weightings from scratch was also a substantial cost (see Equation 4.12). Now a more efficient implementation is outlined, which does not produce mathematically identical estimates, but in fact records a very small performance improvement compared to the results in Section 4.4.

The phase wrapping is automatic when using the 12-bit unsigned integer representation and calculating baseband phase separations by subtraction. Then \( \tilde{u}_k \omega_0 \) is subtracted from the phase separation to set the wrapping range. An efficient method for subsample displacement estimation is interpolation of baseband phase separation. Subsample variation in the weightings seems not to be important, so between pairs of integer shifts the average weighting is used, without interpolation. Thus Equation (C.2b) replaces Equation 4.13 to boost efficiency.

\[
\tilde{u}_{k+1} = \frac{\sum_{y=n\Delta y}^{n\Delta y+Y} W(y; y + U_k, y + U_k + 1) \left\{ (1 - \delta_y) \Delta \phi(y, y + U_k) + \delta_y \Delta \phi(y, y + U_k + 1) \right\}}{\omega_0 \sum_{y=n\Delta y}^{n\Delta y+Y} W(y; y + U_k, y + U_k + 1)} 
\]

(C.2a)

\[
= \frac{\left\{ (1 - \delta_y) \sum_{y=n\Delta y}^{n\Delta y+Y} W(y; y + U_k, y + U_k + 1) \Delta \phi(y, y + U_k) \right\}}{\omega_0 \sum_{y=n\Delta y}^{n\Delta y+Y} W(y; y + U_k, y + U_k + 1)} + \delta_y \sum_{y=n\Delta y}^{n\Delta y+Y} W(y; y + U_k, y + U_k + 1) \Delta \phi(y, y + U_k + 1) 
\]

(C.2b)

In the new expression \( W(y; y + U_k, y + U_k + 1) \) denotes the sum of the weightings at consecutive integer shifts, and \( \Delta \phi(y, y + U_k) \) denotes the correctly wrapped baseband phase separation at shift \( U_k \). For the summation, \( \Delta \phi(y, y + U_k) \) is converted back to floating
point representation by table-lookup, preceded by “bit-compare” to discard any $2n\pi$ offset. Table-lookup is also used to determine the phase weighting part of $\tilde{W}(y; y + U_k, y + U_k + 1)$ during Type A iterations. Once again, Type B iterations are processed efficiently by reusing stored summations.

After displacement estimate convergence, an extra calculation is required at the final subsample shift to estimate location for EPZS with AMC. Table C.1 lists fewer operations for the WPS location estimate, because all necessary summations are accumulated in parallel with Type A iterations to avoid recalculting weightings. As for accuracy indicators, for EPZS the correlation coefficient is calculated, which involves taking the magnitude of the complex correlation that has already been calculated, and dividing through by pre-accumulated squared envelope values to normalise. For WPS, on the other hand, weighted phase separation variance can be calculated more efficiently as an alternative accuracy indicator. The necessary summations are accumulated during Type A iterations, using table-lookup to convert between unsigned integer phase and squared phase in floating point.

C.3 General observations

Generally, both types of algorithm could be accelerated greatly by exploiting parallel processing, although efficient processing on single processor machines is highly desirable. Programme timings indicate that the ratio of pre-processing costs is roughly 4:3 (WPS:EPZS), reflecting the costliness of atan2. In general, the pre-processing cost depends on the form of the raw data, such as if “IQ data” are readily available. Envelope and phase can probably be produced more efficiently for WPS by CORDIC processing, whether in software or hardware [196]. On the other hand, the ratio of deformation estimation costs is roughly 4:5 (WPS:EPZS). All costs scale with the sampling rate. The larger cost is usually that of deformation estimation, which also scales with window spacing and window length. This cost increases further with multi-pass analysis (see Chapter 2), but the effect is small, because estimation should only be repeated for windows that receive a new iteration seed more than a threshold ($e.g., \lambda/4$) away from the previous estimate.
Bibliography


<table>
<thead>
<tr>
<th>Citation</th>
<th>Reference</th>
</tr>
</thead>
</table>


BIBLIOGRAPHY


