STOCHASTIC SIGNAL PROCESSING METHODS FOR SHEAR WAVE IMAGING USING ULTRASOUND

by

Atul Ingle

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Electrical Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

2015

Date of final oral examination: July 30, 2015

The dissertation is approved by the following members of the Final Oral Committee:

Tomy Varghese, Professor, Medical Physics and Electrical Engineering
William Sethares, Professor, Electrical Engineering
John Gubner, Professor, Electrical Engineering
Timothy Hall, Professor, Medical Physics
Christopher Brace, Associate Professor, Biomedical Engineering

To my family.

सर्वे भवन्तु सुखिनः सर्वे सन्तु निरामयाः ।
सर्वे भद्राणि पश्यन्तु मा कश्चित् दुःखमाप्नुयात् ॥

Happiness be unto all, perfect health be unto all.
May all see what is good, may all be free from suffering.

ABSTRACT

Hepatocellular carcinoma is one of the leading causes of cancer related deaths worldwide. Percutaneous needle-based tumor ablation is a minimally invasive treatment method for liver tumors in patients that are not candidates for surgery. In order to prevent recurrence, it is crucial that the right volume of tissue and a safety margin around the tumor are completely ablated during treatment. Although magnetic resonance imaging and computed X-ray tomography are currently used for monitoring the procedure, ultrasound imaging is potentially safer and more cost-effective and can also provide real-time high frame rate imaging capabilities. Information about tissue stiffness obtained using ultrasound elastography can play an important role in locating the boundary of ablated tissue and help in planning and execution of the ablation procedure.

The goal of this dissertation is to develop robust signal processing algorithms to delineate ablation boundaries based on stiffness variations estimated from high frame rate radiofrequency ultrasound echo data. Shear wave velocity is used as a surrogate for tissue stiffness. A method called electrode vibration elastography is used where the ablation needle is vibrated using an actuator and acts as a line source for shear waves. The shear wave pulse is tracked in the imaging plane as it travels laterally away from this line source. Model based denoising algorithms are developed to track the shear wave pulse and reconstruct two dimensional (2D) shear wave velocity images.

In order to generate three dimensional (3D) visualizations of the ablated volume from 2D imaging planes, multiple 2D imaging slices are acquired in a sheaf geometry, with the

ablation needle coinciding with the axis of the sheaf. Shear wave velocity information from the individual 2D imaging planes can be used to reconstruct transverse or C-planes that are stacked to generate a 3D visualization. The reconstruction is performed using a penalized least-squares algorithm. A computationally tractable iterative reconstruction algorithm that can handle 3D grids with millions of points is also developed.

ACKNOWLEDGMENTS

I have many people to thank for helping me directly or indirectly in making this dissertation possible.

I would like to thank my advisor Prof. Tomy Varghese for his advice and support through the uncertainties of graduate school and doctoral research. This dissertation would not have been possible without his constant encouragement and trust in my work. I have benefited immensely from his knowledge of ultrasound elastography and also his advice on research and scientific writing.

I am extremely grateful to Prof. William Sethares for being a source of motivation throughout graduate school and for all the illuminating and entertaining discussions about a plethora of topics ranging from signal processing and mathematics to music and art. His approach toward formulating and solving problems in a variety of application domains has left an indelible mark on the way I think about open-ended research problems.

I would also like to thank Profs. John Gubner, Tim Hall and Chris Brace for serving on my committee, and for their feedback on an early draft of this dissertation. I have benefited from many discussions with Profs. Jim Zagzebski and Jim Bucklew who served on my preliminary examination committee. Prof. Bucklew's academic advice early in my graduate career, and his guidance on topics related to statistical signal processing and Markov chains has been invaluable in shaping the work presented in this dissertation. I thankful to Prof. Ernie Madsen and Gary Frank for help with phantom design and

experimental data acquisition. I am also thankful to Prof. Grace Wahba for guidance on kernel based smoothing methods and for her feedback on the material presented in Chapter 6.

I had a fulfilling internship experience during the two summers spent in Boston working at Philips Ultrasound R&D. I am grateful to Abhay Patil, Scott Dianis and Karl Thiele for sharing their knowledge and expertise in ultrasound imaging algorithms and system design.

I am thankful to many members of the ultrasound group including Kayvan, Nick, Xiao, Chi, JJ, Wenjun, Eenas and Ivan for delightful camaraderie, and especially Ryan DeWall for training me when I first started working with the ultrasound research group. I am very fortunate to have made many wonderful friends in Madison who have helped me survive graduate school.

I cannot thank my parents and brother enough for their unwavering support in all my endeavors, personal and professional. I am very lucky to have such a loving and supporting family back home.

Atul Ingle

July 2015

Madison, WI

TABLE OF CONTENTS

Page

Page

Appendix

Page

Appendix

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Appendix
Figure Page

1

Introduction

> If thou examinest a man having tumors on his breast, thou findest that swelling have spread over his breast; if thou puttest thy hand upon his breast upon these tumors, . . . , they form no fluid, they do not generate secretions of fluid, and they are bulging to thy hand . . .
>
> Edwin Smith Surgical Papyrus, Case 45 (ca. 17^{th} c. BC)

Motivation

Palpation, i.e., physical examination by pushing with the hand or fingers, has been used for locating shallow lesions since times immemorial. Although useful as a first pass, this method is not quantitative. Moreover, manual examination may not reveal stiffness variations due to lesions that are present deep inside the body. Tissue elasticity imaging,

or "elastography," attempts to differentiate between various tissue structures based on local variations in stiffness. In general, an elastography system consists of a method for generating displacements in a region of interest inside the body ("remote palpation") and a system for inferring elasticity properties by measuring these displacements using a medical imaging modality, like magnetic resonance imaging (MRI) or ultrasound.

Elastography The problem of boundary delineation for tumor visualization has been an important signal processing issue in various medical imaging modalities. Since liver tumors may not have sufficient echogenic contrast *vis-á-vis* healthy liver tissue [1, 2, 3], visualizing them on a conventional "B-mode" image is challenging. Ultrasound elastography attempts to derive local mechanical properties of tissue from estimated displacements [4]. It has the potential to augment traditional B-scans and assist the clinician in delineating ablation boundaries more accurately. Unlike X-ray computed tomography (CT) or MRI, traditional ultrasound elastography has been limited to single imaging planes, over which strain is estimated, and the Young's modulus (stiffness) is reconstructed by solving the inverse problem [5, 6, 7, 8]. Shear wave velocity (SWV) and shear modulus can also be estimated for these imaging planes [9]. The accuracy of such methods is limited by the underlying assumptions about tissue elasticity and other geometric and boundary effects.

Liver Cancer Liver cancer is one of the most common forms of cancer in the world and is one of the leading cause of cancer related deaths [10]. Tumor ablation therapy is a minimally invasive procedure that can be used to treat smaller localized tumors. Radiofrequency (RF) and microwave ablation procedures involve insertion of an ablation needle into the affected area and inducing localized heating to thermally coagulate the

cancerous cells. Accurate and rapid visualization of the ablated region has clinical value because it can provide immediate feedback to the clinician about the extent of ablation. Controlling the ablation volume is crucial for preventing recurrence of tumors, arising from the presence of untreated cancerous cells.

The work presented in this dissertation is motivated by the need for robust signal processing methods for ablation boundary delineation as applied to percutaneous needle based liver tumor ablation treatment procedures. The algorithms are primarily designed for electrode vibration elastography (EVE); it uses the ablation needle to generate a shear wave pulse which is tracked in the region of interest using ultrasound [6]. Since ablation is accompanied by stiffening of tissue [11], the speed of this shear wave can be used to demarcate the region that has been treated successfully. With the assumption that tissues are elastic and isotropic and ignoring any high frequency dispersive effects, it is possible to relate shear wave velocity (SWV) c_s and the elastic shear modulus G via the relation:

$$c_s = \sqrt{\frac{G}{\rho}} \qquad [1.1]$$

where ρ is the density of the medium. Ignoring the effects of viscosity, c_s remains constant as a function of frequency. It is also worth noting that the Young's modulus (E) of an incompressible elastic material whose Poisson's ratio is close to 0.5 is related to its shear modulus by

$$G \approx E/3. \qquad [1.2]$$

It is worth noting that these assumptions may fail to hold for anisotropic media. For instance, when estimating mechanical properties of muscle tissue, the SWV along the muscle fibers is usually higher than the velocity in the perpendicular direction [12]. The linear elasticity assumption fails to hold for the case of large displacements because soft

tissue may stiffen as the applied strain increases [13]. Moreover, in viscous tissue, the stiffness varies as a function of frequency [14].

Various algorithms for estimating the velocity of shear waves traveling through disparate media are developed. These algorithms allow two dimensional (2D) mapping of SWVs in the imaging plane. Algorithms for generating three dimensional (3D) reconstructions from parsimoniously acquired data are also developed and analyzed.

Outline

Chapter 2 reviews existing literature about hepatocellular carcinoma, treatment and the role of 2D and 3D ultrasound elastography in ablation procedures and some relevant signal processing literature on 2D and 3D shear wave elastographic reconstruction and boundary delineation.

Chapter 3 discusses electrode vibration elastography, a technique for generating shear waves using an ablation needle, and tracking it using ultrasound. The data acquisition system is described, and methods for beamforming, displacement estimation and wavefront localization are discussed.

Chapter 4 presents various data models for efficient boundary delineation in 2D shear wave elastography. In particular, a Markov model is proposed for shear wave arrival time data in the presence of noise. The model accounts for the *a priori* ignorance about the number and locations of stiffness change points, and practical algorithms with desirable theoretical properties are developed using ideas from maximum a posteriori estimation.

Chapter 5 extends 2D reconstruction algorithms to 3D by reconstructing multiple 2D image planes around the ablation needle using algorithms from Chapter 4 and assembling them into a 3D visualization from a stack of transverse "C-planes."

Chapter 6 develops a kernel based function approximation algorithm using the Matérn kernel which provides a data-adaptive way of choosing the level of smoothness in the final 3D reconstruction. This can be useful for reconstructing the transition region between soft and stiff regions after ablation.

Chapter 7 describes a method for reconstructing the complete 3D volume without resorting to the intermediate step of reconstructing individual transverse planes. A sparse matrix based iterative algorithm is developed using a Markov random field model which enables reconstruction of 3D grids with millions of grid points.

Supplementary material (such as pseudocode for algorithms, mathematical proofs and derivations) is presented in the Appendices.

2

Literature Review

> If I have seen further, it is by standing on the shoulders of giants.
>
> Isaac Newton

2.1 Hepatocellular Carcinoma and its Treatment

Hepatocellular carcinoma (HCC) or liver cancer is one of the leading causes of cancer related deaths in the world with almost 750,000 deaths reported worldwide in 2012 [15]. It is also one of the most common forms of cancer with very high mortality index of 95% [16]. Patients with chronic liver conditions such as hepatitis B, hepatitis C, or cirrhosis have a greater predilection for developing HCC [17]. HCC is the most prevalent type of primary liver cancer. Secondary liver cancer may be caused due to metastasis of cancer from the breast, lung or colon. Other less common types of primary liver cancers include fibrolamellar HCC [18], hepatoblastoma and bile duct carcinoma [19, Ch. 8].

A variety of factors need to be considered before choosing a treatment option for HCC patients. The decision is typically based on the size and number of malignancies and an estimate of overall liver function (such as the Child-Pugh score) [17]. Treatment methods can involve surgical techniques such as resection, transplantation, localized ablation,

or chemotherapeutic methods such as transcatheter arterial chemoembolization (TACE) [20]. For early stage patients, surgical extirpation may be used, especially if the rest of the liver is functioning well. Patients with tumors less than 3–5 cm in diameter may be well suited for localized ablation therapy or ethanol injection to kill the cancerous cells [21]. According to the Barcelona Clinic Liver Cancer staging system, early stage patients are well suited for percutaneous treatments [22].

Localized treatment methods have been an area of active clinical research because they are minimally invasive and can be used for patients that are not candidates for liver transplantation or surgery due to underlying comorbidities. Although randomized clinical trials have not yet provided any conclusive evidence of the superiority of localized techniques over surgical methods or transplantation [23], minimally invasive procedures such as radiofrequency (RF) or microwave ablation have lower risk of post-surgical complications [20, 24]. There is also some clinical evidence for better 3 year survival rates after RF ablation than ethanol injection in for patients with early stage chronic liver disease and tumors smaller than 3 cm in diameter [25].

RF and microwave ablation have a similar principle of operation in that they use electromagnetic energy to heat tissue to over 50°C resulting in cell necrosis. For RF ablation procedures, an RF electrode operating at a frequency of around 500 kHz is inserted directly into the tumor. The electrical conduction circuit is completed through the tissue surrounding the electrode and a grounding pad, and heat is generated as a result of resistive heating accompanying the flow of RF current [26].

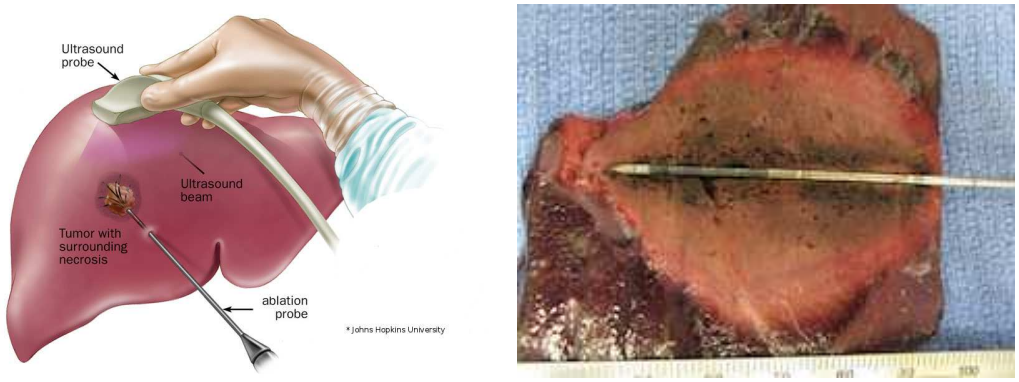
Although microwave ablation also uses electromagnetic energy, it differs from RF ablation in the heat generation mechanism. Microwave ablation antennas usually operate at over 900 MHz [27]. This induces oscillations in the water molecules in the surrounding tissue and some of the energy is dissipated as heat. Since microwave heating does

not depend on the presence of a conductive path for electrons, it has the potential to provide larger and more uniformly treated ablation volumes. However, clinical studies comparing RF and microwave ablation treatment for HCC have not found any difference in survival and recurrence rates [28, 29, 30]. Research has recently shown over 70% survival rates after microwave ablation [31].

2.2 Ultrasound-based Monitoring of Liver Ablation

Two-dimensional (2D) B-mode ultrasound imaging is widely used during liver ablation procedures because of the convenience of low cost, absence of ionizing radiation and real-time operation. The ablation needle appears hyperechoic in a traditional B-mode scan, and hence provides an accurate and cost effective method for guiding needle placement in the tumor. Many commercial ultrasound scanners implement needle guidance modes which use image processing algorithms to emphasize the location of the needle and its tip [32, 33, 34]. Needle mounting brackets attached to the ultrasound transducer ensure that the needle stays in the ultrasound imaging plane.

Tumors, on the other hand, may be difficult to delineate using B-mode images because they may possess similar echogenic properties as background tissue. As a result, stiffness variations may not manifest consistently as echogenicity variations in B-mode images. Liver tumors have been reported to be hypo-, iso- or hyper-echoic in different patients [35, 1]. Ablation reduces the water content of tissue and the ablated region casts a shadow underneath due to increased attenuation in the ablated tissue [36]. Gas bubbles that are generated around the ablation needle appear hyperechoic and introduce artifacts in the B-mode image and also attenuate the ultrasound signals from the distal



(a) (b)

Figure 2.1: (a) B-mode ultrasound imaging is commonly used for needle guidance during percutaneous needle-based ablation procedures. (b) Ablation results in cell necrosis due to protein denaturation and removal of moisture from the treated tissue, which also results in an increase in stiffness.

boundary of the ablation [37]. Hence, although B-mode images may be used to approximately locate the treated volume, they are not reliable for delineation of the ablation boundary.

Ablation of tissue results in protein denaturation and removal of moisture which is accompanied by stiffening relative the surrounding unablated tissue [38]. Therefore ultrasound elastography can provide a more direct method for delineating ablation boundaries. An elastography system consists of three components: (a) a method to apply force and produce displacements, (b) a method to track these displacements, and (c) an algorithm to process the displacement information to infer mechanical properties. Following the pioneering work by Ophir *et al.* [39, 40] on estimating displacement and strain, ultrasound elastography has now evolved into a multitude of imaging techniques that attempt to estimate various mechanical properties such as Young's modulus and viscoelastic parameters.

Magnetic resonance imaging (MRI) can be used for estimating mechanical properties of tissue by estimating displacements. When a harmonic excitation is applied to the medium imaged using an MR scanner, the shear wave displacements get encoded as phase shifts in the nuclear magnetic resonance signal [41]. Magnetic resonance elastography (MRE) [42, 43] can be used to image C-plane sections of the liver and produce shear stiffness maps. However, in the presence of a metallic ablation needle it cannot be used to provide real time feedback about the size of the ablation, unless special MR-compatible ablation probes are used. On the other hand, ultrasound elastography has the potential to provide immediate feedback during ablation allowing clinicians to change ablation parameters in real time.

Ultrasound elastography techniques can be classified based on the method used for generating displacements and mechanical property reconstruction from the measured displacements. Quasi-static strain imaging methods either rely on manual palpation or using the transducer to compress the tissue to be imaged. However, external compression may not be effective for imaging stiffness variations that may be present deep inside the body. For the specific application of needle based ablation procedures, electrode displacement elastography (EDE) is a quasi-static method that uses the ablation needle to generate micron-level displacements that are tracked using ultrasound. This information can be used to generate 2D and 3D strain images of the ablation [44, 45, 46, 47, 46, 48].

In dynamic or transient elastography, an external mechanical device is used to generate displacements that are subsequently tracked using ultrasound. Sandrin *et al.* [49, 50] use an assembly of vibrating rods placed adjacent to the ultrasound transducer to produce shear wavefronts in the image plane. Electrode vibration elastography (EVE) developed by Varghese *et al.* is a transient elastography method well suited for monitoring of percutaneous needle based ablation procedures [6, 7, 51]. The ablation needle is vibrated

using an external actuator and acts as a line source of shear waves. The shear wave pulse travels outward and away from the ablation needle and is tracked using high frame rate ultrasound as it propagates in the imaged medium.

Tissue can be remotely palpated with the ultrasound beam using the acoustic radiation force (ARF) phenomenon. It involves transmission of a short duration high energy focussed ultrasound beam in the region of interest. The resulting pushing force is a function of the intensity of the acoustic wave and the absorption properties of the tissue. In its simplest form, ARF can be used to produce displacements at different locations in the imaging plane and an image of the magnitude of these displacements can be displayed (with the assumption that stiffer areas will undergo a smaller displacement than softer areas) [52]. Local strain images can be obtained by calculating spatial derivatives from this displacement information. An ARF impulse can also be used to generate a shear wave that can be tracked to reconstruct shear wave velocity images [53] and local shear modulus properties [54].

Supersonic shear wave imaging (SSI) generates a line source of shear waves by multiple ARF push pulses along a line to produce a 'Mach cone.' This plane shear wavefront travels perpendicular to the line source [55]. The inverse problem of recovering shear wave velocity and shear modulus from this method has also been studied [56, 9]. Song *et al.* [57] use a "comb push" technique by setting up multiple line sources in the image plane resulting in shear wavefronts traveling in opposite directions which are separately tracked after directional filtering. Crawling wave sonoelastography [58, 59] makes use of two interfering shear waves of slightly different frequencies and uses the interference patterns to derive Young's modulus of the imaged medium.

Another approach aims at mapping viscoelastic properties of tissue by imaging shear waves over a range of frequencies to estimate frequency dependent dispersion of shear

waves together with the Young's modulus [60, 61, 62]. Frequency dependent shear wave velocity measurements can also be made using crawling waves [63] and spatially modulated ultrasound radiation force (SMURF) imaging [64].

There are potentially unlimited configurations and methods for generating shear waves using mechanical excitation depending on the apparatus and the clinical application. The present work focuses on electrode vibration [6, 7] for inducing shear waves with the target application being real-time monitoring of tumour ablation in the liver. An advantage of this method in monitoring liver ablation is that shear waves can be generated *in vivo* by vibrating the same radio frequency-electrode or microwave antenna that is being used for the ablation procedure. In contrast to radiation force type techniques, electrode vibration can be used to obtain larger vibration amplitudes that can be tracked over longer distances in tissue.

2.3 Reconstruction Algorithms for 2D Shear Wave Imaging

Transient shear wave elastography involves tracking of shear wavefronts in the underlying medium using ultrasound displacement estimation techniques. The location of these wavefronts as a function of time can be used for estimating SWV and hence the shear modulus. As is the case with most ultrasound based systems, presence of noise and outliers in unfiltered ultrasound data must be mitigated to attain sufficient signal-to-noise ratios necessary for successful clinical application of this method. The present work proposes various model-based denoising algorithms for SWV reconstruction from noisy ultrasound displacement data. The propagating shear wave consists of a single pulse which is tracked through the imaging plane by recording the time taken for the peak of this pulse to reach different lateral locations in the imaging plane. This process is

repeated at different depths in the imaging plane. This data is referred to as time-to-peak (TTP) data [54] or time-of-flight data [65].

There are various function fitting and denoising methods that have been applied to the problem of filtering noisy TTP information. Palmeri *et al.* [54] apply linear regression followed by statistical goodness of fit criteria. Wang *et al.* [65] apply the random sample consensus (RANSAC) algorithm to address the issue of outliers. They model the TTP curve as a linear function of the spatial coordinates with the coefficients as free parameters to be estimated. The RANSAC algorithm proceeds by randomly drawing subsets of the full dataset. Next, it calculates the parameters of the hypothesized linear model using a least-squares fit and then identifies and removes possible outliers at each iteration. A specific cost function is used to gauge the quality of fit.

Rouze *et al.* [66] proposed using the Radon sum transform for estimating SWV in a single medium. Using a three dimensional map of lateral location, time and displacement the algorithm extracts a trajectory in the lateral location vs. time plane that gives the maximum sum of displacements. This trajectory is assumed to be the path of propagation of the shear wave. Intuitively, this method gives the best fit line along the locations of the displacement peaks thereby smoothing out the effect of noise.

In a preliminary study, Bharat *et al.* [6] discuss the phenomenon of change in the slope of the TTP profile when a shear wave travels through an interface between different media. A least-squares fit is applied to the noisy TTP data prior to calculating the slope of the curve at various lateral locations for estimating SWV and locating any slope change points. Various algorithms to automatically and reliably detect these slope change locations of the TTP curve are developed in this dissertation. TTP slope change detection is an important problem as it indicates the presence of a transition boundary between regions of different stiffnesses.

As an alternative to the TTP approach, McLaughlin *et al.* [9] use correlation based pattern matching to locate a shear wave pulse of a known shape at different locations in the medium. The issue of noise smoothing is handled implicitly by use of the Eikonal equation thereby avoiding derivatives of noisy data and circumventing the issue of solving an ill-posed inverse problem. In another recent paper, Klein *et al.* [67] present an improved correlation based method for estimation of shear wave arrival times by using a penalized optimization procedure to account for the phenomenon of pulse shape broadening.

In a more recent paper, Zhao *et al.* [68] analyze the effect of ultrasound imaging system parameters such as transducer, frequency and imaging depth on SWV estimates obtained using acoustic radiation force. They use a novel plane wave ultrasound insonification technique and software beamforming which allows much higher frame rates than traditional focused ultrasound imaging. The SWV reconstruction algorithm applies a least-squares fit to the TTP measurements.

It is also possible to calculate SWV by directly inverting the wave equation [69]. This method involves taking higher order time and spatial derivatives of displacement estimates obtained from any standard ultrasound-based motion tracking algorithm. However, this method is fraught with noise, notwithstanding the use of standard noise reduction techniques such as median filtering for removing outliers and mean filtering for smoothing.

A theoretical analysis of the complete signal processing chain employed in a standard ultrasound based SWV imaging system was presented in a paper by Deffieux *et al.* [70]. Using methods from classical estimation theory, they derive a Cramér-Rao lower bound on the variance of any unbiased estimator of shear modulus when the shear wave propagates in a homogeneous linearly elastic medium with no interfaces. Tracking the

propagation of a shear wavefront through multiple media is more challenging due to uncertainty regarding wave velocities within different media and exact locations of interfaces. As a result, direct application of statistical function fitting techniques provides no theoretical guarantees on detecting the interfaces and slopes accurately. Viewed from the point of view of TTP data, this problem is equivalent to fitting a continuous piecewise linear function with unknown slopes, unknown breakpoints and unknown number of segments.

The need for piecewise linear regression arises in many different fields, as diverse as biology, geology, and the social sciences. The algorithms developed here are general enough to be applicable to other fields, although the main application of interest is ultrasound shear wave elastography. In many real world applications, the local slope values of an observed noisy function have interesting physical interpretations. Most of the existing methods do not directly address slope estimation; rather, they attempt to fit a model to the data. Instead, directly estimating the slopes and breakpoints is useful when the slopes correspond to the variables of interest and the breakpoints correspond to where those variables change.

The topic of slope estimation from noisy data is quite old; an early paper can be traced back to 1964 where the popular Savitzky-Golay differentiator [71] was introduced. Their main idea is to use a locally windowed least squares fit to estimate the slope at each data sample, where the window coefficients are chosen to satisfy a certain frequency response that mimics a high pass filter together with some level of noise averaging. Another similar technique that is used in statistics is called locally weighted least squares regression (LOWESS) [72]. However, these methods undesirably smooth out the breakpoint locations in when data has sharp transitions or jumps. In contrast, model-based techniques

can be used to preserve sharp transitions (for instance, by explicitly allowing for sharp slope transitions using a Markov model).

In some situations, the raw data can be massaged using a preprocessing step so that it becomes piecewise linear. The simplest example is the case of piecewise constant data — the running sum (integral) of such data yields a piecewise linear function. Ratkovic and Eng [73] discuss a statistical spline fitting approach combined with the Bayesian information criterion (BIC) to detect abrupt transitions in political approval ratings. As a special case, their method can be applied when function values stay almost constant over long intervals and occasionally shift to a new value. In another application, Bai and Perron [74] use statistical regression techniques to detect multiple regime shifts in interest rate data.

Closely related problems of piecewise linear regression for noisy data have been addressed over the years. For example, an early paper by Hudson [75] focuses on a technique to obtain piecewise linear fits in a least-squares sense with only two segments. The break point location is included as a parameter in the least squares optimization problem. The method is extended by dealing with multiple breakpoint locations on a case-by-case basis which becomes combinatorially intractable as the number of breakpoints increases. Bellman [76] suggests a dynamic programming approach when the number of breakpoints is unknown. However, this method requires the use of a large number of grid points for accurate results. Gallant and Fuller [77] generalize this to the fitting of arbitrary polynomials with unknown breakpoints while requiring the function to be composed of segments with continuous first derivatives. They apply a non-linear optimization routine (Gauss-Newton minimization) to fine-tune breakpoint locations while minimizing the squared error relative to the data. Another non-parametric approach involves use of edge preserving penalized optimization such as total variation

minimization [78]. Denison *et al.* [79] use a Markov chain Monte Carlo approach to fit piecewise polynomials with different numbers and locations of knot points. Tishler and Frey [80] discuss a maximum likelihood approach to fit a convex piecewise linear function expressed as a point-wise maximum of a collection of affine functions with unknown coefficients. Maximum likelihood estimates are obtained by running a constrained optimization routine for a smoothed approximation of a mean squared error (MSE) cost function to bypass non-differentiability issues. A similar data model coupled with data clustering heuristics is utilized in a more recent paper by Magnani and Boyd [81] on fitting convex piecewise linear functions.

The use of adaptive methods is an attractive way of handling the issue of unknown number of breakpoints. One of the first algorithms using this technique was proposed by Friedman [82] under the name "adaptive regression splines (ARES)." Recursive partitioning is used to obtain better partitions of the set of data points at each iteration. Either goodness of fit criteria or generalized cross validation is used to estimate the number of partitions. On similar lines, Kolaczyk and Nowak [83] apply the method of recursive dyadic partitioning and fit a smooth function in each partition using maximum likelihood estimation. A penalty term for the number of partitions is introduced to trade off model complexity and quality of fit. In recent work, Saucier and Audet [84] propose a different class of adaptively constructed basis functions that can capture the transition points in otherwise piecewise smooth functions.

In [85] Bai and Perron discuss the problem of detecting structural changes in data without requiring the estimated function to be piecewise linear or even continuous. Their related paper [74] discusses a dynamic programming approach to obtain a least sum of squares fit. The model order is determined by using the Akaike information criterion (AIC) [86] and they impose a minimum limit on the "run length" of each segment in

the piecewise model. In contrast, the present work proposes a dynamic program that generates optimum maximum a posteriori (MAP) estimates based on a stochastic finite state hidden Markov model.

In the signal processing literature, two kinds of paradigms have been applied to this problem — Bayesian estimation and pattern recognition approaches. Punskaya *et al.* [87] model the function using the number and locations of the breakpoints as free parameters with certain prior distributions. The posterior density of the parameters conditioned on the noisy data is estimated through Monte Carlo techniques. In response to this method, Fearnhead [88] proposes a direct method for estimating parameters of the same model without resorting to Monte Carlo simulations and exploiting a Markov property in the model that allows calculation of the probability of future data points conditioned on the most recent breakpoint location.

2.4 Regularized Reconstruction Methods in 3D Elastography

Two dimensional (2D) ultrasound has been widely applied to tissue stiffness measurements in ablation monitoring procedures in the liver [47, 46, 48, 89]. 2D electrode vibration shear wave imaging method can be extended to three dimensions (3D) by utilizing RF echo signals acquired over a "sheaf" of imaging planes. A sheaf is defined as a collection of planes that intersect along a common axis. This method of acquisition is naturally suited to electrode vibration elastography where shear wavefronts travel outward from the vibrating needle resulting in a cylindrical wavefront, with the needle acting as a line source.

There has been growing interest in 3D ultrasound imaging and elastography; one evidence being the evolution of literature on this topic in the last two decades. Elliott [90] notes the increasing use of 3D data acquisition among ultrasound sonographers

to circumvent the limitations of traditional 2D ultrasound. Various authors have analyzed reconstruction algorithms for 3D B-mode imaging [91, 92] of different anatomical structures using a variety of transducer types and scanning arrangements. Quasistatic freehand elastography has received much research attention [93, 94]. These 2D elastography techniques can be naturally extended to 3D in various ways. Freehand elastography can be performed by manually translating the transducer probe through parallel imaging planes. Alternatively, a "wobbler" that mechanically rotates or translates an array transducer using a stepper motor may be used. Other authors have proposed using more elaborate robotic techniques [95] for accurate control of the location of the transducer in 3D space. Each image plane is processed using standard elastography algorithms and a 3D rendering is generated [96]. Freehand elastography can be augmented with accurate optical [97] or magnetic position sensors that precisely record the coordinates of the transducer. Fortunately, for tumor ablation monitoring using EVE, the ablation needle provides a good reference for manually aligning the imaging plane. Elaborate tracking and registration systems have the potential to improve reconstruction accuracy. The present work uses only a crude alignment strategy relying on the assumptions that the underlying 3D structure is fairly symmetric about the needle axis, independent sheaves (with relatively small misregistration errors) are acquired, and the final 3D reconstructions are averaged. This averaging operation can be used to reduce misregistration errors, assuming these errors are small and zero mean.

Results on 3D quasi-static strain [45, 98] and transient SWV reconstruction [99] for prostate imaging have been reported in literature. These methods typically utilize a 1D array transducer to acquire data over multiple image planes. For instance, a wobbler can be used to obtain elevational slices, or a rotating probe can be used to obtain angularly spaced azimuthal slices. Strain images can be reconstructed on each image plane using

standard quasi-static elastogram reconstruction techniques, and the individual planes can be stacked together to produce a 3D rendering of the strain. Lee *et al.* [100] have reported significant improvement in detection of cancerous breast lesions when B-mode imaging is augmented with freehand 3D shear wave imaging. Literature on full 3D reconstruction of SWVs and shear stiffness is still quite nascent. In recent work by Wang *et al.* [12], 3D reconstruction of muscle fiber orientation was achieved by mapping group and phase velocities of the shear wave wave set up using acoustic radiation force. Although the use of matrix transducer arrays for volume ultrasound imaging is gathering pace, linear and curvilinear array transducers are still the most widely used transducer types. Therefore, the ability to generate volume rendering akin to CT or MRI using 2D ultrasound data has clinical value [90].

3

Electrode Vibration Elastography

> "... we have described a new, quantitative method for imaging the elastic properties of compliant tissue. ... The method could become useful in a number of applications."
>
> J. Ophir, *et al.*, Elastography: a method for imaging the elasticity of biological tissues

3.1 Introduction

The raison d'être of an elastography system is to create an image of mechanical elasticity properties of the underlying medium. Some of the earliest work in ultrasound elastography involved development of tissue displacement estimation methods using ultrasound echo data with the help of cross-correlation based or Doppler based techniques [101, 102, 103]. The opening quote of this chapter is from an early paper on ultrasound elastography [39] which demonstrated the use of ultrasonic echo data to produce a map of tissue strain in the image plane. Modern elastography algorithms are now capable

of estimating various mechanical properties that include strain, Young's modulus, shear wave velocity (SWV), shear modulus, and viscoelastic parameters. These properties are estimated from local tissue displacements and an assumed model for tissue elasticity. An image of elasticity properties is called an "elastogram." In general, an ultrasound elastography system consists of three distinct components:

1. a method for generating displacements,

2. a method for tracking these displacements across frames, and,

3. an algorithm for estimating stiffness properties from these displacements.

The third step is not necessary if the desired elastogram only consists of displacement information.

Vibration of the ablation needle is a unique method applicable to the case of ablation monitoring where the needle is already in place for treating the tumor. The technique is therefore called electrode vibration elastography (EVE). This chapter describes the experimental setup used for EVE which includes a vibrating actuator that is operated in synchronization with the ultrasound scanner. The design and use of tissue-mimicking (TM) phantoms used for experimental validation of the algorithms developed in this dissertation is also described. Since shear waves can travel at quite high speeds, it necessitates the use of high frame rate ultrasound acquisition to track these waves in the imaged medium. One method involves phase-locked acquisition which uses multiple vibrations to assemble "pseudo-high frame rate" data. The second method uses plane acoustic insonifications to bypass the low frame rate limitation of traditional focused transmit acquisitions.

A cross-correlation based displacement estimation algorithm is used in the second step to obtain a displacement movie across all frames. In its simplest form, this algorithm calculates axial displacements (in the direction of ultrasound beamlines, also called A-lines) by cross-correlating segments of radiofrequency (RF) echo data in a small neighborhood along the A-line. 2D displacement tracking can be used to correct for lateral displacements by running the cross-correlation algorithm along the beamlines and also laterally across A-lines.

A shear wave pulse can be tracked in both space in time with the knowledge of particle displacements in each frame. Ignoring any dispersive effects due to viscoelasticity, the wave equation (Helmholtz equation) can be inverted to calculate shear wave velocity [69]. This method is difficult to use in practice because it requires calculation of spatial and temporal second order derivatives from noisy displacement data. In EVE, it is assumed that the shear wave pulse propagates purely laterally and the wave velocity can be estimated from times of arrival at different locations away from the needle.

This chapter describes the three components of an EVE system — the data acquisition setup which produces the shear wave perturbation and captures RF echo data, the displacement estimation routine which calculates frame-to-frame displacements, and the assumed shear wave propagation model used for the various image reconstruction algorithms developed in subsequent chapters.

3.2 Experimental Setup

3.2.1 Tissue-mimicking Phantoms

The phantom based study involved data acquisition from two TM phantoms with similar design but slightly different mechanical properties[96]. Both phantoms consist

of a centrally situated stiff ellipsoid in a softer background material which simulates the presence of a tumor in cirrhotic liver tissue. The stiff ellipsoid is intended to model an ablated region. An irregular tumor structure whose stiffness was in between the ablated region and the background material was attached on one side of the ellipsoid. The phantom material is composed of a dispersion of microscopic oil droplets in a gelatinous matrix. The difference in stiffness is mainly achieved by controlling the proportion of oil in the material; a detailed discussion about the design and properties of such phantom materials is given in the paper by Madsen *et al.* [104]. A stainless steel rod was bonded to the center of the ellipsoid in order to mimic the role of a radio-frequency electrode or a microwave antenna in an actual ablation procedure. This rod was used for generating shear waves in the phantom with the help of an actuator. A representative cross-sectional view of the structure of both phantoms is shown in Fig. 3.1. The gelatin block was 14 cm$\times$14 cm$\times$9 cm. This block was placed in an open top 1 cm thick acrylic container. A layer of safflower oil about 2 cm deep was poured over the top surface to prevent desiccation.

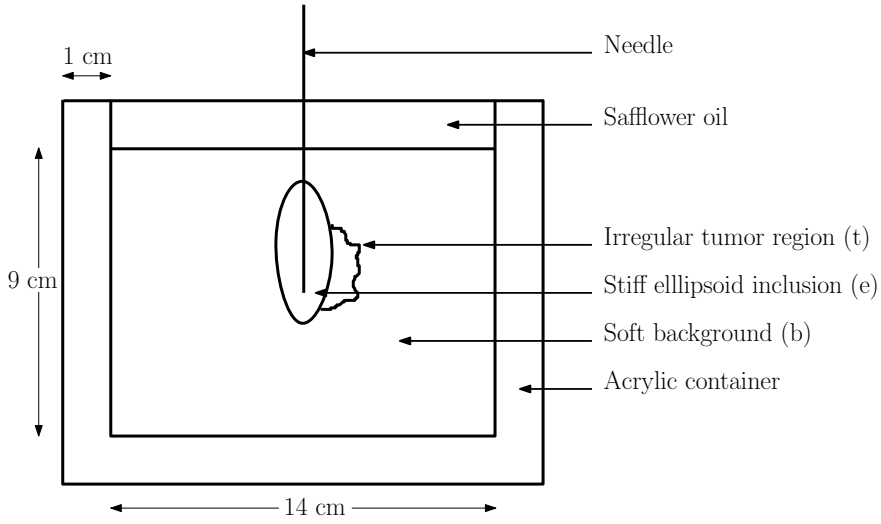


Figure 3.1: A schematic representation of the cross-sectional view of the two phantoms. This diagram is not to scale.

3.2.2 Data Acquisition Setup

Ultrasound echo data was acquired using an Ultrasonix SonixTouch machine (Ultrasonix Medical Corporation, Richmond, BC, Canada) and a software tool developed using the Ultrasonix Software Development Kit [7]. The 9L4 linear array transducer operated at a frequency of 5 MHz was used for obtaining ultrasound RF echo data. The transducer has a 6 dB bandwidth of 33%, a transmit F-number of 2.6 an acoustic pulse duration equal to 1 transmit cycle of the operating frequency of 5 MHz. The effective line density was equal to the number of elements i.e. 128 lines over a lateral extent of 3.8 cm.

Shear wave pulses are generated by vibrating the needle using a piezoelectric actuator (Physik Instrumente (PI) GmbH, Karlsruhe, Germany). The actuator produces a half-sine wave pulse of 100 μm amplitude and 30 ms width. In order to synchronize data collection and pulse generation, the actuator controller is set up to trigger the ultrasound scanner and the actuator simultaneously. A schematic view of an EVE setup is shown in Fig. 3.2. A pulse deformation applied to the needle sets up a shear wave pulse where the wave source is a line coinciding with the needle and the shear wavefronts travel cylindrically outward and away from this line. Ignoring any viscoelastic attenuation, the amplitude of this cylindrical wavefront decays as the square root of the distance away from the needle.

3.3 Displacement Tracking

3.3.1 Sequential Tracking

Since the frame rates with traditional B-mode ultrasound imaging are not sufficient to track a shear wave pulse (which travels with a velocity of a few meters per second), a

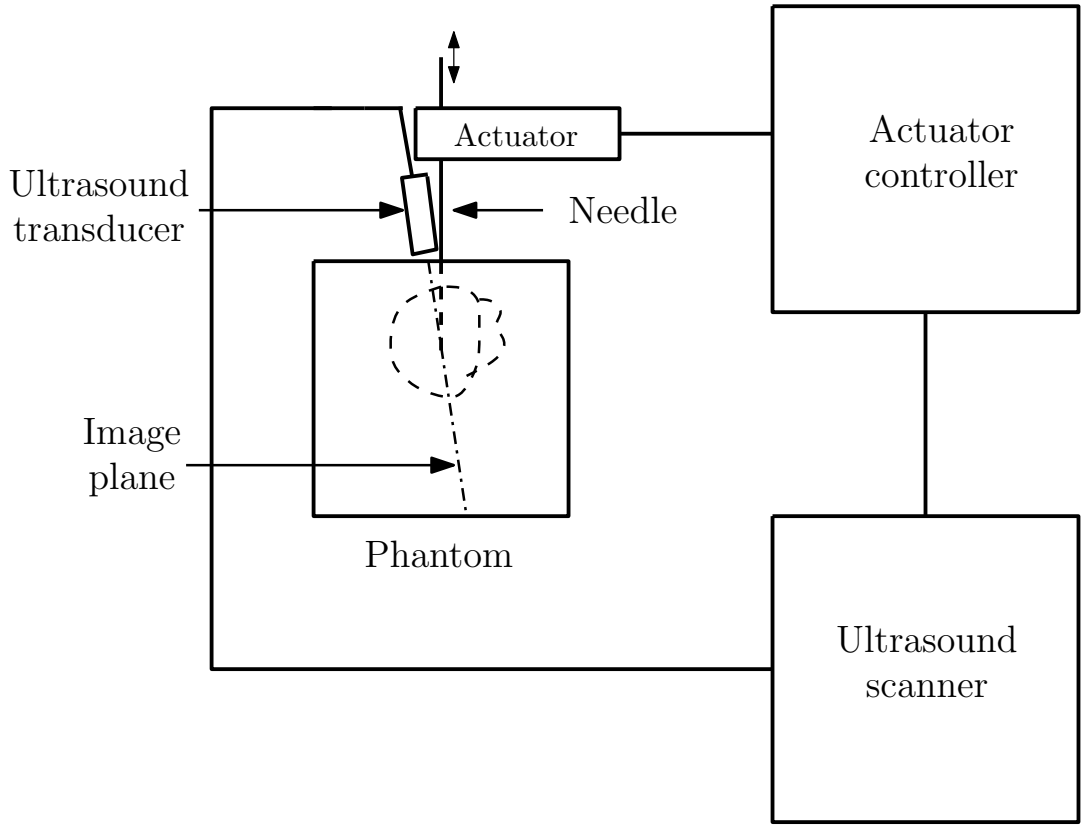


Figure 3.2: This schematic shows the data acquisition system used in electrode vibration elastography experiments. The needle is vibrated using an actuator. Data acquisitions are phase-locked by synchoronizing the ultrasound scanner pulse sequencing with the actuator motion. The image plane is adjusted in such a way that it contains the needle.

sequential tracking technique developed by DeWall and Varghese [7] can be used. This method is similar to the phase locking technique used in MR elastography [105]. In this method, the needle is vibrated multiple times and with each vibration a different location at a specified lateral distance away from the needle is scanned using the ultrasound system (Ultrasonix SonixTouch, Richmond, BC, Canada). Vertical strips of RF ultrasound echo data are then registered and assembled to obtain an apparent high frame rate frame sequence over the entire image plane. The phase-locked acquisition scheme assumes that the needle is vibrated identically in each cycle and the image plane is not changed during the acquisition. For *in vivo* application, additional motion filtering and synchronization

may be needed to account for patient motion. Alternatively, a high frame rate plane wave imaging technique that does not require multiple perturbations can be used.

3.3.2 Plane Wave Tracking

High frame rate plane wave ultrasound imaging [106] can be used to scan the entire image plane in a single transmit. The idea is similar to synthetic aperture imaging [107], however all the transducer elements are activated simultanously to produce a plane acoustic wavefront, instead of firing each element individually to generate spherical wavefronts. Plane wave imaging operates by setting the focal depth to infinity and using an aperture size equal to the number of elements in the linear transducer array. In the absence of steering, this is equivalent to applying zero delays to the transducer elements. As soon as the transmit sequence is completed, the ultrasound scanner receives parallel channel data simultaneously for each transducer element bypassing the hardware beamformer.

In general, the plane wave can be transmitted at an arbitrary angle α as shown in Fig. 3.3. Let x and z denote the lateral and axial dimensions respectively with the origin of the coordinate system is fixed at the left edge of the transducer. Let d denote the distance between consecutive transducer elements and consider the acoustic energy received at element # n after scattering from a particle at location (x,z) in the image plane.

The acoustic energy scattered by this particle at location (x,z) is received at different elements of the array with different delays depending on the total path length and the speed of sound. The beamformer must coherently sum up the contributions from this particle by applying appropriate delays to the pre-beamformed scan lines for each receive element of the transducer. Let $s_n(t)$ denote the raw prebeamformed channel

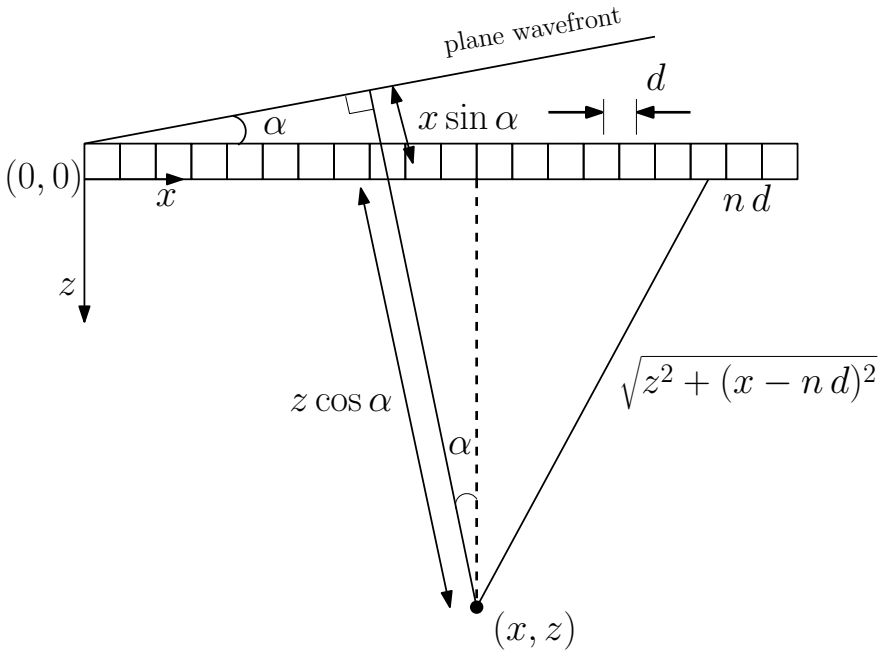


Figure 3.3: Angular insonification with plane waves requires some geometrical calculations for beamforming. The total path length can be expressed in terms of the angle α and the coordinates of the arbitrary pixel (x, z).

waveform received at element # n in the array. Then the beamformed signal for location (x, z) in the image plane is given by:

$$s(x, z) = \sum_{n=-a/2}^{a/2} w_n s_n(\tau_\alpha(x, z, n))$$

where a denotes the aperture over which the summation occurs, w_n is a window function and $\tau_\alpha(x, z, n)$ is the total time delay for the acoustic wave to travel from the transducer to the particle at (x, z) and its reception at element # n. Using the geometry depicted in Fig. 3.3, the time delay is given by:

$$\tau_\alpha(x, z, n) = \frac{1}{v_s}\left((x \sin \alpha + z \cos \alpha) + \sqrt{z^2 + (x - n\,d)^2}\right)$$

where v_s is the speed of sound (assumed 1540 m/s).

Since, in practice, the beamformer must process sampled data, all the variables in the beamforming equations are discretized as follows:

$$n \in \{0, 1, 2, \ldots, 127\}$$
$$x \in \{0, d, 2d, \ldots, 127d\}$$
$$z \in \{0, 1, 2, \ldots, \#\text{ axial samples}\}.$$

The time delay τ is rounded off to the nearest integer and used to index into the pre-beamformed channel data s_n.

The beamformed RF data frames are averaged to generate the final compounded image.

3.4 Shear Wave Pulse Propagation Model

The shear wave pulse vibration results in a displacement field which depends on both the spatial coordinates and time t. Assuming that the ablation needle acts as a line source of plane shear wave fronts traveling laterally (in the x direction) in the image plane, with purely axial displacements u_z, the wave equation can be written as:

$$\mu \frac{\partial^2}{\partial x^2} u_z(x,t) = \rho \frac{\partial^2}{\partial t^2} u_z(x,t). \qquad [3.1]$$

Here μ is the shear modulus of the medium and ρ is the density [108], and the equation suggests a constant shear wave velocity of $\sqrt{\frac{\mu}{\rho}}$. When using a linear array transducer to image the medium, all measurements are restricted to the $x - z$ plane.

Although the shear wave velocity can be calculated directly by inverting the wave equation it is seldom used in practice because it requires taking the second spatial and

time derivatives of noisy displacement measurements. Instead, the "time of arrival" of a shear wavefront propagating purely in the lateral x direction in a homogeneous medium can instead be used to calculate SWV from the Eikonal equation [108, Thm. 2.3]:

$$\left|\frac{d}{dx}T_a(x)\right| = \frac{1}{c_s} \tag{3.2}$$

where the time of arrival $T_a(x) = \arg\inf_{t>0}|u_z(x,t)| > 0$, i.e. the first time instant when the displacement at a given location x is non-zero.

In practice, due to the noisy nature of displacement estimates obtained from ultrasound data, using the definition of $T_a(x)$ directly in an algorithm is not advisable. Typically, the practical approach involves using more information such as the peak of the wavefront which may be easier to estimate from noisy data [7, 54]. Alternatively, the shape of the wave packet can be used in as part of a cross-correlation [109, 9] or an elastic matching [67] algorithm. For the data acquired from the experimental setup described in Section 3.2, the location of the pulse peak is estimated with subsample precision by fitting a quadratic polynomial using five samples around the peak. The following proposition suggests that under certain idealizing assumptions about the wave propagation, using the time of arrival calculated using the wave peak also satisfies (3.2).

Proposition 3.1 Assume a half-sine wave shaped shear wave pulse traveling with a velocity c_s in a homogeneous medium with no diffraction and attenuation. The displacements can be described by $u_z(x,t) = u_z(x-ct) = A\cos\left(\frac{\pi}{W}(x-ct)\right)$ for $-W/2 \leq x - ct \leq W/2$ and 0 otherwise, where A is the amplitude and W is the pulse width. Then the time of peak $T_p(x) = \arg\max_{t>0}|u_z(x,t)|$ satisfies (3.2).

Proof: In the absence of pulse shape broadening due to diffraction (i.e., ignoring viscoelastic effects),

$$\begin{aligned} T_p(x) &= \arg\max_{t>0} |u_z(x,t)| \\ &= \arg\max_{t>0} \ \mathbf{1}_{[-\frac{W}{2},\frac{W}{2}]}(x-ct) A \cos\left(\frac{\pi}{W}(x-ct)\right) \\ &= \arg\inf_{t>0} \ \mathbf{1}_{[-\frac{W}{2},\frac{W}{2}]}(x-ct) A \cos\left(\frac{\pi}{W}(x-ct)\right) + \frac{W}{2} \\ &= T_a(x) + \frac{W}{2}, \end{aligned}$$

where $\mathbf{1}$ is the "indicator function" defined as:

$$\mathbf{1}_A(t) = \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{if } t \notin A. \end{cases} \qquad [3.3]$$

For future reference, the complement of this function is defined as $\mathbf{1}_A^c(t) := 1 - \mathbf{1}_A(t)$.

This implies that $\frac{d}{dx} T_p(x) = \frac{d}{dx} T_a(x)$ and hence $T_p(x)$ also satisfies (3.2). □

SWV reconstructions are obtained for each scan plane using the time to peak algorithm [54]. A shear wave pulse is produced by vibrating the needle using an actuator, and multiple frames of RF ultrasound data are acquired simultaneously. The scan sequence and reconstruction algorithm is discussed in more detail in the paper by DeWall *et al.* [7]. Displacements between consecutive frames are estimated using a 1D cross-correlation algorithm and the peak displacements at each pixel are recorded to localize the shear wave pulse in space and time. The reciprocal of the slope of a plot of the time of arrival of the shear wave pulse as a function of lateral distance away from the needle is used to estimate shear wave velocity at each pixel in the scan plane [110]. As shown in Fig. 3.4, the line source of shear waves is assumed to coincide with the needle, and the

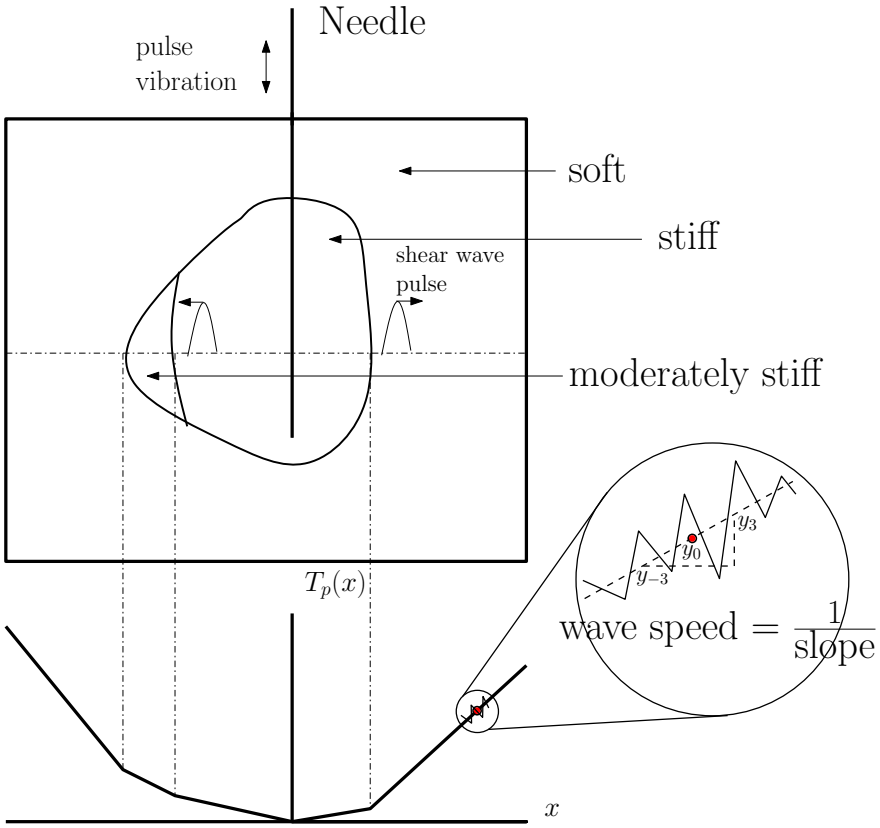


Figure 3.4: This figure shows an idealized time-to-peak (TTP) plot, i.e. a plot of $T_p(x)$ vs. lateral distance x at a fixed depth z for an EVE experiment. The local slope of this plot is equal to the reciprocal of the SWV and the slope change points correspond to the stiffness boundaries.

wave pulse is tracked on either side of the needle along lines of constant depth to create an arrival time plot of $T_p(x)$ vs. x.

3.5 Results

Although a single plane wave insonification can provide ultrafast frame rates of over 10,000 frames/s, single plane wave transmit is seldom sufficient to obtain good quality RF data useful for displacement estimation. Images created from a single plane wave

transmit possess much lower lateral resolution than traditional focussed transmit ultrasound images. The presence of hyperechogenic regions in the imaging plane causes a "hazy" appearance in laterally adjacent regions, as seen in Fig. 3.5. Similar artifacts were reported previously in a paper by Montaldo *et al.* [111].

When such RF data frames are used in a displacement estimation routine, the poor lateral resolution causes ambiguity in the location of the shear wave pulse. An axial displacement movie from an electrode vibration TM phantom experiment is shown in Fig. 3.6. A shear wave pulse can be clearly visualized in the top half of the image plane traveling from left to right, however, the location of the peak of the wave is difficult to follow in the lower half of the image. The lower half of the image contains "false" displacements that bleed laterally from the motion of the hyperechoic region.

Fig. 3.7 shows B-mode images of a point target phantom imaged using angular compounding method. Observe that the haze that extends laterally from the point targets is significantly reduced when multiple angular transmits are compounded. Angular insonifications reduce the imaging frame rate by a factor equal to the number of angles used. This, however, does not pose a serious drawback in practice because a frame rate of around 1000 frames/s is sufficient to track a shear wave pulse.

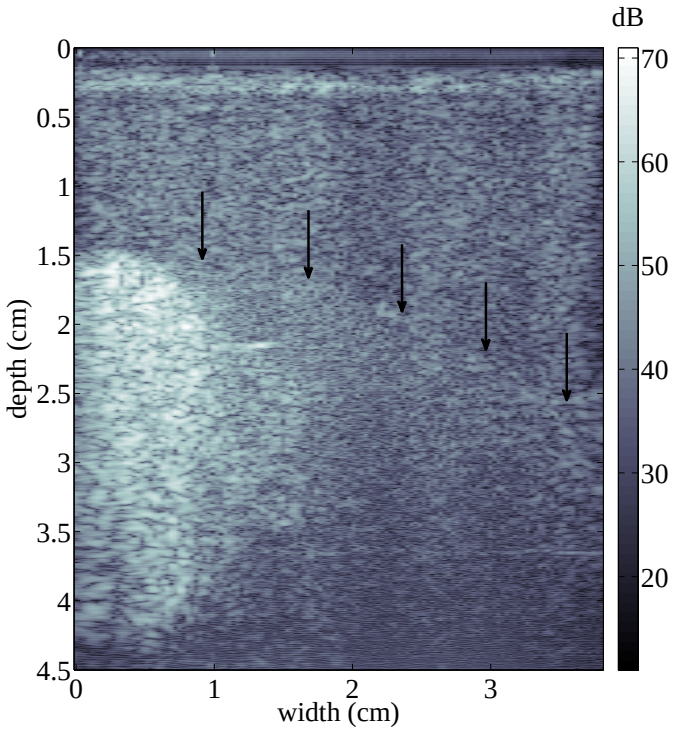


Figure 3.5: When using a single plane wave insonification to image a hyperechoic region, the bright areas in the image bleed into the hypoechoic regions due to the lateral spread of the point spread function. This causes a hazy appearance in the B-mode image outlined using the arrows. The effect on displacement estimates is shown in Fig. 3.6.

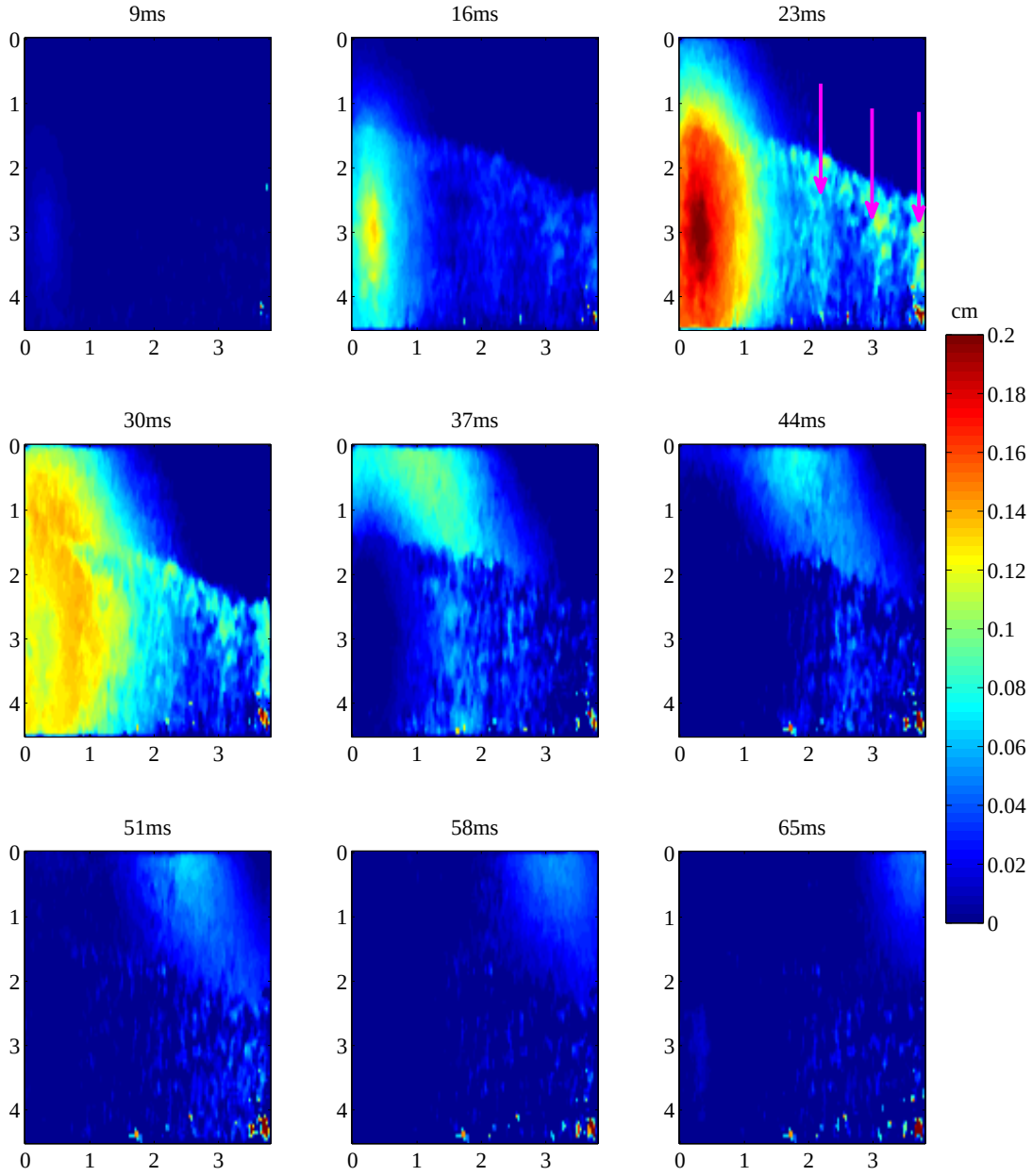


Figure 3.6: This movie of axial displacements shows false displacement peaks that prevent reliable tracking of the shear wave pulse in locations laterally adjacent to the hyperechoic region. Note that the false displacement region (shown with the pink arrows) is correlated with the hazy region in the B-mode image from Fig. 3.5. The image plane consists of a stiff inclusion in the higher echogenic region surrounded by softer background material.

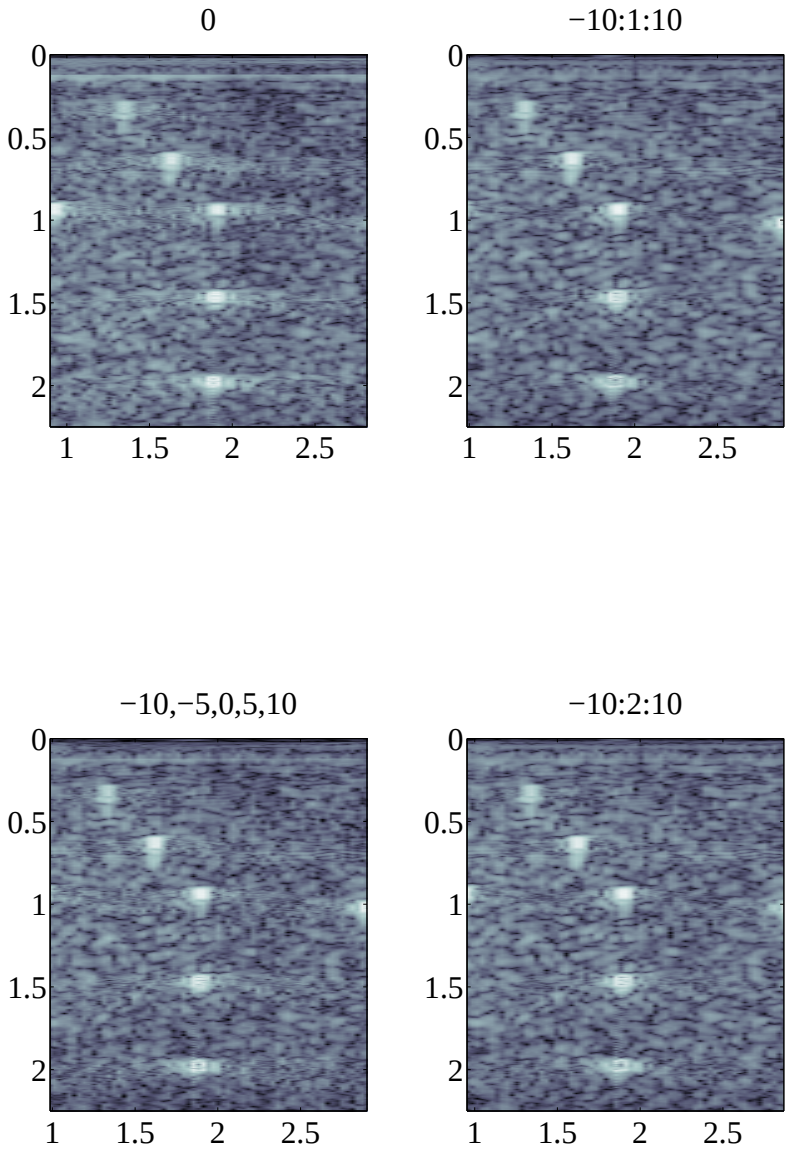


Figure 3.7: The lateral spread of point targets in a point target phantom is significantly reduced with plane wave compounding as seen in these images. The subplot titles denote which angles are compounded to produce the final image.

4

Markov Model for Shear Wave Tracking

One should not regret the past. One should not worry about the future. Wise men act by the present time.

Sanskrit proverb

A Markov chain is a random process with the property that, conditioned on its present value, the future is independent of the past.

G.R. Grimmett, D.R. Stirzaker,
Probability and Random Processes

4.1 Introduction

One of the main assumptions used for generating shear wave velocity (SWV) images from high frame rate displacement data is that of purely lateral propagation. This assumption rests on the fact that the particle displacements produced after needle vibration are perfectly parallel to the direction of vibration of the needle. Hence (at least

in the region where the needle is bonded to the surrounding tissue), it is reasonable to assume that the wave motion is similar to an ideal transverse wave. This means that the shear wave pulse travels laterally along the face of the transducer, while all particle displacements are axial.

This chapter proposes various model based statistical signal processing algorithms for solving this inverse problem of inferring SWV from information about wave arrival times. The arrival times are estimated by tracing the time-to-peak (TTP) displacement as a function of distance away from the needle along lines of constant depth. However, in practice, the measured TTP data can be quite noisy, and some form of noise suppression must be applied before finite differencing can be used to obtain SWVs. Moreover, the wave may travel through different stiffness interfaces causing abrupt changes in SWV. It is important to detect these changes because in the ablation application they correspond to boundaries between treated (ablated) and unablated tissue.

The first model uses a "brute force" approach by describing the TTP data as a piecewise linear function with an unknown number and locations of breakpoints. Supposing the number of breakpoints is known, the problem is reduced to a multi-parameter non-convex optimization problem. A model order selection criterion is used to choose the most parsimonious model i.e. the piecewise linear fit that explains the noisy data with the fewest breakpoints possible.

The second model presented here encodes the ignorance about the number and locations of breakpoints into a hidden Markov model framework where the local slope values (i.e. SWVs) play the role of hidden states. A particle filter is designed to unravel the most likely sequence of slopes from noisy data. Since particle filtering can be computationally burdensome, a second approximate algorithm that uses a discrete set of slope values is proposed.

Finally, these algorithms are validated synthetic data, data from finite element simulations and experimental data obtained from tissue-mimicking (TM) phantoms.

4.2 Piecewise Linear Fitting Problem

The TTP data obtained over each image plane can be used to estimate SWV by calculating the reciprocal of the slope of the TTP curves at different depths. However, some form of smoothing must be applied prior to calculating the slope because differentiating noisy data will amplify noise. An important consideration for this noise filtering algorithm is to preserve edge details between the stiff inclusion and the soft background. An optimization algorithm that fits a continuous piecewise linear function to the noisy TTP curves [112] is designed with a function model parametrized by the locations of breakpoints and slopes of individual segments. Let the number of segments in the fit be denoted by B, the unknown breakpoint locations by $\{\lambda_i\}_{i=0}^{B}$ and unknown segment slopes by $\{m_i\}_{i=1}^{B}$. A piecewise linear function is modeled as:

$$f(x) = \sum_{i=0}^{B-1} \mathbf{1}_{[\lambda_i, \lambda_{i+1})}(x) \left[m_{i+1}(x - \lambda_i) + \sum_{j=1}^{i} m_j(\lambda_j - \lambda_{j-1}) \right] \qquad [4.1]$$

where $\mathbf{1}$ denotes the indicator function defined in (3.3).

A specific case of the piecewise linear function model with $B = 3$ is shown in Fig. 4.1. It is assumed that the noisy TTP data, $\{y_i\}_{i=1}^{N}$, originates from this true underlying function. Therefore a constrained least-squares problem can be formulated to estimate the free parameters in this model:

$$\begin{aligned} &\text{minimize} && \frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i))^2 \\ &\text{subject to} && m_i > 0, \text{ and } \lambda_0 < \lambda_1 < \ldots \lambda_B. \end{aligned}$$

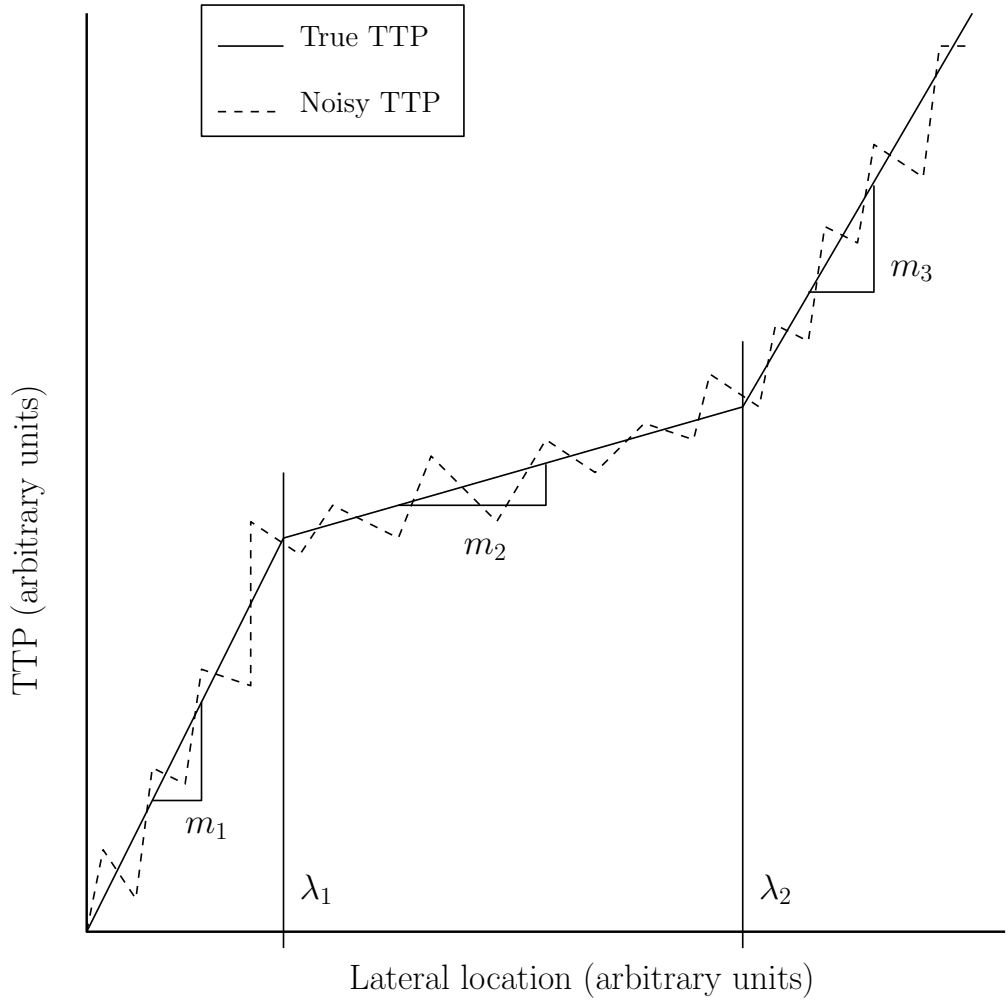


Figure 4.1: A piecewise linear model with unknown breakpoints (λ_1, λ_2) is used for estimating local slope values (m_1, m_2, m_3). The reciprocals of these slope values are used as estimates of local shear wave velocities at different lateral locations away from the electrode.

The least-squares optimization problem is solved using a standard sequential quadratic programming [113] numerical optimization routine. Other methods such as interior point optimization, log-barrier algorithms, or stochastic optimization methods such as simulated annealing [114] can also be used.

In reality, the number of segments B is not known *a priori*. Therefore the algorithm must choose a reasonable B automatically. This problem of "model order selection"

is handled using the Akaike information criterion (AIC) [86] which trades off mean-squared error and model complexity by minimizing the following function by choice of B:

$$\mathsf{AIC} = N \log(\mathsf{MSE}/N) + (2B - 1)$$

where MSE is the residual mean squared error between the data and the fit. The intuition behind using this criterion is that it not only penalizes a bad fit (large MSE) but also penalizes the number of segments (model complexity), thereby providing a safeguard from over-fitting. The optimization problem is solved with $1 \leq B \leq 10$ and the best B is chosen to minimize AIC.

4.3 Hidden Markov Model

A shear wave can be assumed to travel with a constant speed as long as it propagates in the same homogeneous medium, but the speed changes when it enters a dissimilar medium via an interface. Therefore, to a first approximation, an ideal TTP curve should appear to grow linearly with a constant slope, except at interfaces, where the slope should abruptly change to a new value. Ultrasound imaging at high frame rate can be used for tracking axial tissue displacements as a function of time from which the TTP can be computed for various lateral locations. These signal processing steps leading up to the TTP estimates introduce some noise, which will be assumed to have a Gaussian distribution. A stochastic hidden Markov model can be formulated to represent the physical phenomenon of the traveling shear wave together with this noisy TTP measurement.

The following notation is used to describe the hidden Markov model. Let

Z_n be the actual (noiseless) value of the TTP curve at a lateral location n

X_n be the noisy measured value of the TTP curve

w_n be i.i.d. Gaussian noise with density $\mathcal{N}(0, \sigma^2)$.

S_n be the most recent slope value of the curve

U_n be a continuous valued uniformly distributed random variable

M_n be a two dimensional "state variable," (Z_n, S_n)

B_n be a Bernoulli random variable with $\mathsf{p}(B_n = 0) = p$ and $\mathsf{p}(B_n = 1) = 1 - p$ for some $0 < p < 1$.

Individual components of a vector are denoted using superscripts. For instance, $R^{(l)}$ is used to denote the l^{th} component of a vector R. The shorthand notation $R_{l:l+m}$ is used to denote the set of vectors $\{R_l, R_{l+1}, \ldots, R_{l+m}\}$. It is assumed that the TTP data is obtained at equally spaced sampling points along the lateral dimension. Hence it is possible to normalize the distances to 1 unit per sample.

The relationship between these random variables can now be summarized as follows:

$$Z_{n+1} = Z_n + S_n \qquad [4.2]$$

where

$$S_{n+1} = \begin{cases} S_n & \text{if } B_n = 0 \\ U_n & \text{if } B_n = 1 \end{cases}$$

and

$$X_n = Z_n + w_n \qquad [4.3]$$

for $n \geq 0$, with the initial distribution $S_0 \sim U_0$.

These equations model the presence of interfaces in a probabilistic manner. At any lateral location the shear wave continues to propagate in the same medium with a probability p or encounters an interface with a probability $1-p$ and the slope switches to a new values chosen uniformly randomly over some reasonable interval $[c_l, c_u]$ of minimum and maximum possible shear wave speeds. In most practical scenarios, the value of p can be chosen to be close to 1 because one expects to propagate through very few interfaces.

A hidden Markov model belongs to a more general class of dynamical models that can be characterized by a "state evolution equation" and an "observation equation." In their standard forms, the state evolution equation relates the current state to its immediate predecessor and the observation equation relates the output of the system to its current state. The stochastic model that generates the noisy TTP data can be cast into this standard form as follows:

$$M_{n+1} = \begin{bmatrix} M_n^{(1)} + M_n^{(2)} \\ (1-B_n)M_n^{(2)} + B_n U_n \end{bmatrix} \qquad [4.4]$$

and

$$X_n = M_n^{(1)} + w_n. \qquad [4.5]$$

The dependence between various random variables involved in this model are pictorially represented in Fig. 4.2.

4.4 Particle Filter

The goal of a filtering algorithm is to unravel the most likely state values given a set of observed data, that is, to provide a good estimate for $\{M_1, \ldots, M_N\}$ given $\{X_1, \ldots, X_N\}$,

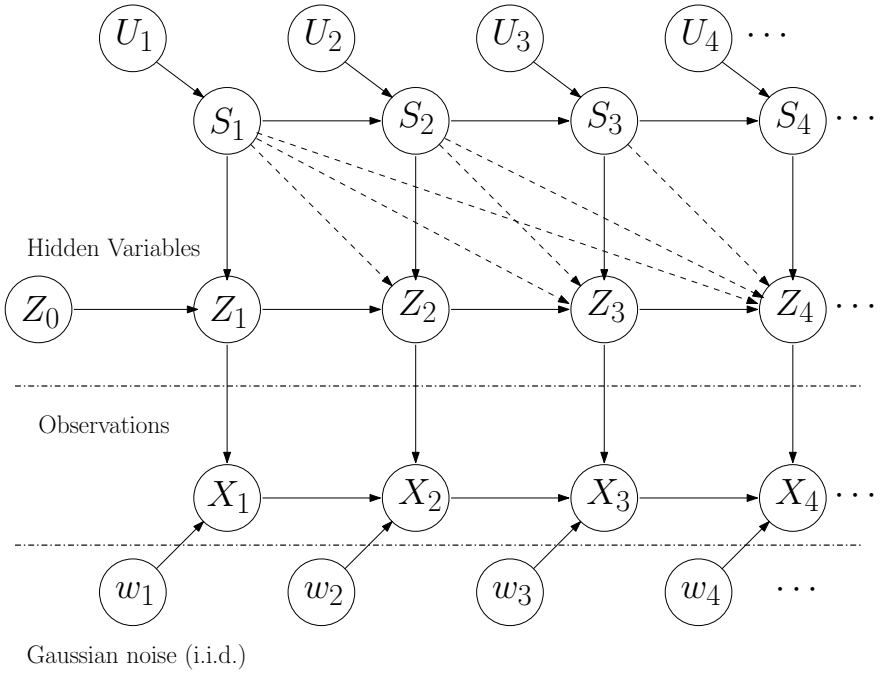


Figure 4.2: A pictorial representation of the relationships between various random variables (input, hidden states and output) of the hidden Markov model is shown here.

where N is the number of data points. When the state space is continuous valued, one of the ways of solving this problem is through particle filtering. A brief discussion of the concept of particle filtering is included here for completeness. Detailed description of various embellishments in the basic algorithm and implementation issues are discussed in a tutorial article by Arulampalam *et al.* [115]. Some practical implementation issues and simulation results with real world examples are discussed in the paper by Gilks and Berzuini [116]. Development of the theory behind optimality of particle filters for hidden state estimation when the noise distribution is not Gaussian can be found in the paper by Bergman *et al.*[117].

Particle filtering is a Monte Carlo technique that calculates the approximate probability density function of the state variables conditioned on observed data. This is achieved by traversing all the data points sequentially and updating the density estimate based

on an application of Bayes' rule. Particle filtering methods have been applied in the context of quasistatic elastography [118] and for sequential tracking problems in medical imaging [119, 120]. Bayesian methods have been previously used for motion tracking [121, 122] using RF ultrasound data.

Suppose that the algorithm is at the k^{th} data point out of the N available observations. The joint density of all the states up to the k^{th} data point can be expressed approximately as a weighted sum of dirac-delta functions as:

$$p(M_{0:k}|X_{1:k}) \approx \sum_{i=1}^{N_s} \omega_k^i \delta(M_{0:k} - M_{0:k}^i) \qquad [4.6]$$

where N_s is the number of points used in the discrete approximation, $\{M_k^i\}_{i=1}^{N_s}$ is a set of random points in the state space, and $\{\omega_k^i\}_{i=1}^{N_s}$ is a set of corresponding weights that satisfy $\sum_{i=1}^{N_s} \omega_k^i = 1$. The number of points N_s is typically quite large ($\sim 10^3 - 10^4$) so that the discrete approximation is close to the actual continuous density function.

For the dirac-delta approximation to hold, each random point M_k^i must be drawn according to the density function $p(M_{0:k}|X_{1:k})$. However, since this probability density function is unknown, a special technique of importance sampling [115, 116] is used to generate these points. The points $\{M_{0:k}^i\}_{i=1}^{N_s}$ can instead be generated from a different "importance sampling density" function $q(M_{0:k}|X_{1:k})$. The dirac-delta approximation will continue to hold provided the weights $\{\omega_k^i\}_{i=1}^{N_s}$ are normalized proportional to the ratio $\frac{p(M_{0:k}^i|X_{1:k})}{q(M_{0:k}^i|X_{1:k})}$. The advantage of this scheme is that the sampling density q can be chosen arbitrarily. Suppose that the sampling density is factorizable as $q(M_{0:k}|X_{1:k}) = q(M_k|M_{0:k-1}, X_{1:k})q(M_{0:k-1}|X_{1:k-1})$. Then the new sample set can be obtained by simply appending to the old set $\{M_{0:k-1}^i\}_{i=1}^{N_s}$ a new set of random points $\{M_k^i\}$ for the current state. These new points can be drawn i.i.d. from the density $q(M_k|M_{0:k-1}, X_{1:k})$. Next,

using Bayes' rule, the weight-update equation becomes:

$$\omega_k^i = \omega_{k-1}^i \frac{p(X_k|M_k^i)p(M_k^i|M_{k-1}^i)}{q(M_k^i|M_{0:k-1}^i, X_k)}. \quad [4.7]$$

Moreover, since the main goal is to estimate only the current state, the dirac-delta approximation can be modified to:

$$p(M_k|X_{1:k}) \approx \sum_{i=1}^{N_s} \omega_k^i \delta(M_k - M_{0:k}^i) \quad [4.8]$$

by marginalizing over the previous states. A convenient choice for the sampling density is:
$q(M_k^i|M_{0:k-1}^i, X_k) = p(M_k^i|M_{k-1}^i)$ which leads to the weight update equation given by:

$$\omega_k^i = \omega_{k-1}^i p(X_k|M_k^i). \quad [4.9]$$

This choice of the sampling density makes computational implementation tractable. The state evolution and observation equations are both accessible and are used for generating the random samples and updating the weights, respectively.

In the course of generating new state samples from the old ones and updating the weights, particle filters face the issue of degeneracy. This is a phenomenon where most of the sample points stray into the low probability density region of the density function thereby carrying almost zero weight. This prevents an accurate representation of the high density areas of the density function. A procedure called "resampling" can be applied to bypass this issue [115]. Resampling generates a modified particle set from a degenerate particle set in such a way that a majority of the sample points remain concentrated in high probability density regions.

As a further fallout of the resampling procedure, in the worst case all the particles tend to collapse to a single point and there is a loss of diversity in the set of samples that are being used for approximating the density function. To counteract this, a regularized particle filter implementation is used in this paper where a random perturbation is added to each point after resampling. Under certain assumptions [123], the optimal choice for these perturbations is the Epanechnikov kernel. This choice follows from results on kernel density estimation which show that this kernel generates a minimum variance estimate of the density function [124].

The expected value of the current state can be obtained from the density estimate as:

$$\mathbf{E}[M_k|X_{1:k}] = \sum_{i=1}^{N_s} \omega_k^i M_k^i. \qquad [4.10]$$

Moreover, the variance of this estimate can also be calculated:

$$\mathbf{Var}[M_k|X_{1:k}] = \mathbf{E}[(M_k)^2|X_{1:k}] - (\mathbf{E}[M_k|X_{1:k}])^2 \qquad [4.11]$$

where the square is taken elementwise. This may be of limited value in practice because clinicians are seldom interested in variance at a fixed point and are more interested in the distribution of values on a larger region of interest (ROI).

The sequential filtering procedure outlined above does not exploit the fact that data is available for batch processing for the ultrasound shear wave tracking problem. Better smoothing estimates may be obtained if the observations from the future are also incorporated into the current state estimation step. This can be done using a lookahead-window smoothing method that waits for L samples into the future before generating

the current state estimate as

$$\mathbf{E}[M_k|Y_{1:k+L}] = \sum_{i=1}^{N_s} \omega_{k+L}^i M_k^i. \qquad [4.12]$$

Choosing L to be very large has pitfalls of oversmoothing [125] and wasting some data samples towards the end of the dataset. As a further improvement, Doucet *et al.* also propose the "fixed lag smoothing" algorithm [125] which runs a backward smoothing step to update the weights according to:

$$\hat{\omega}_k^i = \sum_{j=1}^{N_s} \hat{\omega}_{k+1}^j \frac{\omega_k^i \; p(M_{k+1}^j|M_k^i)}{\sum_{l=1}^{N_s} \omega_k^l \; p(M_{k+1}^j|M_k^l)}$$

for $k = n-1, \ldots, 1$ and $\hat{\omega}_n^i = \omega_n^i$. Note that this is computationally more burdensome to implement in real-time applications (because of the need to process and store N_s^2 different values of $p(M_{k+1}^j|M_k^l)$ at each k). Therefore this last equation is only used in evaluation of the HMM approach in simulations; the lookahead-window smoother is used for the experimental data. The pseudocode for these algorithms is listed in Appendix B.

4.5 Discrete State Approximation

An implicit assumption in the Markov model developed here is that the slopes can take on values on a closed bounded interval $[c_l, c_u]$ with upper and lower limits $0 < c_l < c_u < \infty$ known *a priori*. For instance, in the ultrasound-based wave tracking application, the values of c_l and c_u can be obtained from the underlying physics which dictates that such mechanical waves travel with speeds between 0.5 to 10 m/s in homogeneous tissue. With the knowledge of c_l and c_u, the given data vector can be translated and rescaled

so that all slope values lie in the interval $[0,1]$. Hence, without loss of generality, it suffices to design a slope estimation algorithm that operates with a finite set of slopes $\mathcal{S} = \{0, \frac{1}{(M-1)}, \ldots, \frac{M-2}{M-1}, 1\}$. Intuitively, this quantization step is justified because in the presence of noise it is impossible to detect the difference between slope values that differ only slightly.

4.5.1 Maximum a Posteriori (MAP) Estimation

The probabilistic structure of the data generation process can be expressed using a conditional density function of the unknown function values conditioned on the observed data vector. Let $\mathsf{p}_{Z_{0:N}|X_{1:N},\theta}(Z_{0:N} = z_{0:N}|X_{1:N} = x_{1:N}, \theta)$ be the posterior probability density of the function values $z_{0:N}$ conditioned on the observed data points $x_{1:N}$. The goal of the inference problem is to unravel the most likely sequence of hidden states (slope values $s_{1:N}$) that produced the observed function values $x_{1:N}$. This can be posed as a MAP estimation problem where the posterior density of the unknown states conditioned on the observed data is maximized.

Proposition 4.1 Suppose

$$\begin{aligned} Z_0 &= 0 \text{ with probability 1,} \\ Z_n &= Z_{n-1} + S_n, \\ X_n &= Z_n + w_n \end{aligned} \qquad [4.13]$$

for $n = 1, \ldots, N$ where w_n are i.i.d $\mathcal{N}(0, \sigma^2)$. A Markov structure is imposed as follows:

$$S_n = \begin{cases} S_{n-1} & \text{with probability } p \\ U_n & \text{with probability } 1-p \end{cases}$$

for $n = 2, \ldots, N$ where $U_n \sim \mathcal{U}(\{0, \frac{1}{(M-1)}, \ldots, \frac{M-2}{M-1}, 1\} \setminus \{S_{n-1}\})$ denotes a discrete uniform random variable taking on one of $M-1$ possible slope values and the initial slope value is drawn uniformly as $S_1 \sim \mathcal{U}(\{0, \frac{1}{(M-1)}, \ldots, \frac{M-2}{M-1}, 1\})$. Then the states Z_n form a HMM second-order Markov chain.

Proof: Using (4.13), it is easy to switch between the random variable Z_n and S_n by invoking the recursive relationship $Z_n = Z_{n-1} + S_n$. The first two states are handled as special cases. Note that $\mathsf{p}(Z_0 = z_0) := 1$ for $z_0 \equiv 0$, and $\mathsf{p}(Z_1 = z_1 | Z_0 = z_0) = \mathsf{p}(S_1 = z_1) = 1/M$ for $z_1 \in \mathcal{S}$ and 0 otherwise. Also,

$$\begin{aligned}
\mathsf{p}(Z_2 = z_2 | Z_1 = z_1, Z_0 = z_0) &= \mathsf{p}(S_1 + S_2 = z_2 | S_1 = s_1) \\
&= \mathsf{p}(S_2 = z_2 - z_1 | S_1 = z_1 - 0) \\
&= \mathsf{p}(S_2 = z_2 - z_1 | S_1 = z_1 - z_0)
\end{aligned}$$

which depends only on z_2, z_1 and z_0, since $Z_0 = 0$ with probability 1. In general, for $3 \leq n \leq N$,

$$\begin{aligned}
\mathsf{p}(Z_n = z_n | Z_{0:n-1} = z_{0:n-1}) &= \mathsf{p}(Z_{n-1} + S_n = z_n | Z_{0:n-1} = z_{0:n-1}) \\
&= \mathsf{p}(z_{n-1} + S_n = z_n | Z_{0:n-1} = z_{0:n-1}, Z_{n-1} - Z_{n-2} = z_{n-1} - z_{n-2}) \\
&= \mathsf{p}(S_n = z_n - z_{n-1} | Z_{0:n-1} = z_{0:n-1}, S_{n-1} = z_{n-1} - z_{n-2}) \\
&= \mathsf{p}(S_n = z_n - z_{n-1} | S_{n-1} = z_{n-1} - z_{n-2})
\end{aligned}$$

which only depends on z_n, z_{n-1} and z_{n-2}. Hence Z_n is a second-order Markov chain. The transition probabilities can be written explicitly as:

$$\mathsf{p}(Z_n = z_n | Z_{n-1} = z_{n-1}, Z_{n-2} = z_{n-2}) = \begin{cases} p & \text{for } z_n = 2z_{n-1} - z_{n-2} \\ 0 & \text{for } z_n < z_{n-1} \\ \dfrac{1-p}{M-1} & \text{otherwise} \end{cases} \quad [4.14]$$

for $3 \leq n \leq N$. □

The posterior density can be further simplified using Proposition 4.1 and Bayes' theorem as follows:

$$\begin{aligned} \mathsf{p}(z_{0:N} | x_{1:N}, \theta) &= \frac{\mathsf{p}(x_{1:N} | z_{0:N}, \theta) \mathsf{p}(z_{0:N} | \theta)}{\mathsf{p}(x_{1:N} | \theta)} \\ &= \frac{\mathsf{p}(x_1 | z_1, \theta) \mathsf{p}(z_1 | z_0, \theta)}{\mathsf{p}(x_{1:N} | \theta)} \prod_{j=2}^{N} \left[\mathsf{p}(x_j | z_j, \theta) \cdot \mathsf{p}(z_j | z_{j-1}, z_{j-2}, \theta) \right]. \end{aligned} \quad [4.15]$$

For MAP estimation, it convenient to work with the log posterior density derived using (4.14) and (4.15):

$$\begin{aligned} \log \mathsf{p}(z_{0:N} | x_{1:N}, \theta) &= -\log \mathsf{p}(x_{1:N} | \theta) - \frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - z_n)^2 \\ &\quad + \sum_{n=2}^{N} \log \left[p \, \mathbf{1}_{\{0\}}(s_n - s_{n-1}) + \frac{1-p}{M-1} \mathbf{1}^c_{\{0\}}(s_n - s_{n-1}) \right]. \end{aligned} \quad [4.16]$$

where $\mathbf{1}_A(t)$ is the indicator function for the subscripted set A, as defined previously in (3.3).

In a standard HMM, this maximization problem is efficiently solved using a dynamic program called the Viterbi algorithm [126]. For the present model, the standard Viterbi approach cannot be applied because of the second-order transition structure coupled

with the fact that the state space of the hidden state Z_n changes with n. Two algorithms to solve the MAP estimation problem for the present model are shown in this section, including a computationally tractable dynamic program discussed in Section 4.5.1.1.

The MAP optimal slope sequence can be obtained by maximizing (4.16) as a function of the slope sequence, which is equivalent to the following optimization problem:

$$s^*_{1:N} = \arg\max_{s_{1:N}} \Bigg(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - z_n)^2 + \sum_{n=2}^{N}\log\Bigg[p\,\mathbf{1}_{\{0\}}(s_n - s_{n-1}) + \frac{1-p}{M-1}\mathbf{1}^c_{\{0\}}(s_n - s_{n-1})\Bigg]\Bigg). \quad [4.17]$$

The term containing $\mathsf{p}(x_{1:N}|\theta)$ is dropped because it does not depend on $s_{1:N}$. Besides producing the MAP optimal solution, the objective function has certain intuitively appealing characteristics. The first summation on the right hand side of (4.17) is just the sum squared error, which must be minimized. The second summation acts like a penalty term that encourages longer runs of constant slope. In general, MAP estimation problems are non-trivial owing to the presence of multiple local maximizers and computational complexity associated with optimization of multi-variable functions. Fortunately, the present maximization problem can be handled efficiently by taking a piecemeal approach furnished by the dynamic programming principle [127]. The following algorithm differs from the standard forward-backward algorithm used with HMMs which only handles one level of hidden states.

4.5.1.1 Dynamic Program

The objective function in (4.17) is a sum of individual terms that depend only on z_n, s_n and s_{n-1}. One way of visualizing a maximization algorithm is using the trellis in Fig. 4.3 where each depth n shows the possible values k that the state Z_n can realize.

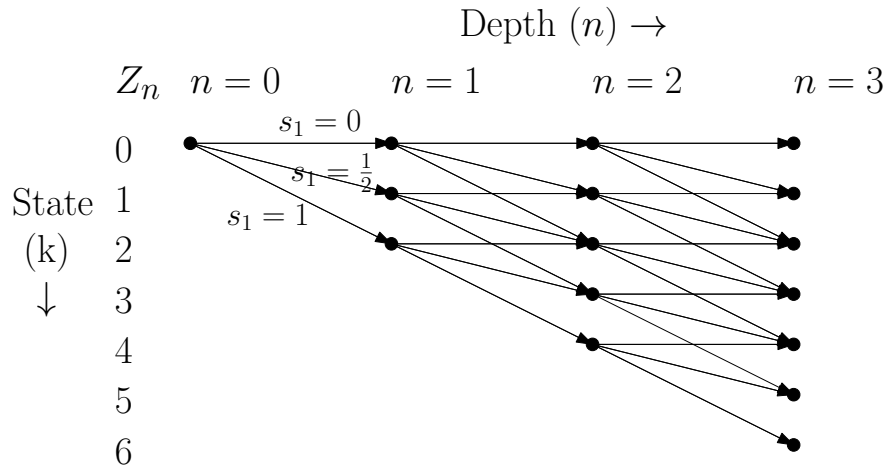


Figure 4.3: A representative diagram of a trellis structure used for finding the MAP optimal slope sequence. This trellis is shown with slope values quantized into three bins and a data vector length of 4.

Each branch has an associated branch "reward" which corresponds to individual terms of the summation in (4.17). The peculiarity about this trellis is that each branch has a variable reward depending on the previous branch chosen along the optimizing path. A dynamic program called FastTrellis is shown in Fig. B.4. This algorithm works by storing the optimum path vector $\mathbf{\Pi}(n,k)$ and its associated reward $I(n,k)$ coming into each node k at every depth n of the trellis. The path that maximizes the partial sums in (4.17) is chosen as the optimal path for each node. For each $1 \leq n \leq N$, the path vectors $\mathbf{\Pi}(n,k)$ and cumulative rewards $I(n,k)$ are updated by appending new nodes to the optimal paths terminating at nodes at the previous depth $n-1$.

Note that the branch rewards referred to in the innermost loop of FastTrellis are composed of two terms: the first term is the negative squared error with respect to the data and the second term is an additional reward for maintaining the same slope value as the previous data sample. The final output of this algorithm contains the best path reaching the deepest level in the trellis and the value of the maximized sum shown in (4.17). Referring back to Proposition 4.1, it is also worth noting that the pseudocode

shown in Fig. B.4 implicitly converts the second order Markov structure on the Z_n process into a first order Markov process by keeping track of a pair of values, viz., the recent slope value $s_n = z_n - z_{n-1}$ and the previous slope value $s_{n-1} = z_{n-1} - z_{n-2}$, when deciding the branch cost.

It is instructive to compare the computational complexity of this procedure *vis-á-vis* a standard Viterbi algorithm that operates on a constant height trellis. The basic unit of computation is assumed to consist of two floating point additions, one floating point multiply and one compare — these operations are needed for calculating the rewards for each branch in the trellis in Fig. 4.3. The terms $\log p$ and $\log\left(\frac{1-p}{M-1}\right)$ can be precomputed, hence do not enter the complexity analysis. The height of the trellis at any depth n is $n(M-1)+1$. Therefore the worst case number of computations at depth n is of the order $\mathcal{O}(n^2M^2)$. Summing over a total trellis depth of N levels, the worst case computational complexity of FastTrellis is $\mathcal{O}(N^3M^2)$. Although this is worse than the $\mathcal{O}(NM^2)$ complexity [126] of a standard Viterbi algorithm with a constant trellis height, it is still an improvement over a brute force approach which will require searching over M^N slope value sequences.

4.5.1.2 Smooth Optimization

The indicator functions in (4.17) cause the objective function to be non-differentiable. This can be addressed by approximating an indicator using a narrow Gaussian pulse:

$$\mathbf{1}_{\{0\}}(t) \approx \exp\left(-\frac{t^2}{\alpha^2}\right)$$

where α controls the width of the roll-off. As $\alpha \searrow 0$ a narrow spike of height 1 is obtained at $t = 0$ thus approximating the indicator function with increasing precision. After discarding terms that do not depend on $s_{1:N}$, the following approximation to the

optimization problem in (4.17) is obtained:

$$s^*_{1:N} = \arg\max_{s_{1:N}} \Bigg(-\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left(x_n - \sum_{j=1}^{n} s_j \right)^2 + \sum_{n=2}^{N} \log \left[p\, e^{-\frac{(s_n - s_{n-1})^2}{\alpha^2}} + \frac{1-p}{M-1} \left(1 - e^{-\frac{(s_n - s_{n-1})^2}{\alpha^2}} \right) \right] \Bigg). \quad [4.18]$$

This smooth approximation of the original problem can be solved using standard constrained optimization routines such as gradient descent [128, 113]. Although it does not guarantee that the slope values lie in the discrete set $\mathcal{S}$, this issue can be bypassed either by rounding the slopes to the nearest quantization bins, or by using integer-programming to solve the optimization problem. Algorithmic complexity of this method varies based on the optimization algorithm used. In general, for a gradient descent type approach that constructs a Hessian matrix of size $N \times N$, the worst case complexity of inverting the Hessian matrix is $\mathcal{O}(N^3)$. This is similar to the complexity of the dynamic programming method. It should, however, be noted that FastTrellis produces an exact solution, whereas there is no guarantee that the optimization method will converge to $s^*_{1:N} \in \mathcal{S}^N$.

4.5.2 Alternating Maximization Algorithm for Estimating Model Parameters

Choosing the model parameters (σ^2, p) is crucial to ensure a sensible fit to the noisy data and enable numerical evaluation of the log posterior probabilities. Unfortunately, these parameters that define the stochastic model are seldom known in advance. Although reasonable values can be guessed by manually preprocessing the data, an automatic method for choosing these parameter values is desirable. The Baum-Welch algorithm which is a special case of expectation-maximization (EM) algorithm is a standard

method for parameter estimation [129, 130, 131] in HMMs. In general, it is difficult to prove that the EM procedure converges unless certain assumptions about the likelihood function are made, unimodality being a common assumption [132]. Other approaches include approximate EM iterations via Monte Carlo integration to approximate the expectation step [133].

Although it is easy to set up the EM iteration equations for the present model, it is computationally intractable for larger values of M and N. A derivation can be found in Section C.1. The following result provides a useful alternative re-estimation procedure that bypasses the computational issues with the full EM approach. Unlike the EM approach which ascends the observed data likelihood function, the MAPSlope algorithm ascends the complete data likelihood function by alternately maximizing with respect to the slope sequence and the model parameters. So, it has a similar advantage as an EM algorithm in that it increases the likelihood at each iteration. Under an additional minor technical assumption, it can also be shown that the procedure converges.

Theorem 4.2 Let $s_{1:N}$ be the current slope sequence estimate, $s^*_{1:N}$ be the new sequence estimate obtained by running FastTrellis with the current model parameter values $\theta = (\sigma^2, p)$. Let $\theta^* = (\sigma^{*2}, p^*)$ be re-estimated from the new slope sequence estimate as:

$$\sigma^{*2} = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \sum_{j=1}^{n} s^*_j\right)^2 \qquad [4.19]$$

and

$$p^* = \frac{1}{N-1}\sum_{n=1}^{N-1} \mathbf{1}_{\{0\}}(s^*_{n+1} - s^*_n). \qquad [4.20]$$

Then the complete data likelihood function satisfies

$$\mathsf{p}(x_{1:N}, s_{1:N}|\theta) \leq \mathsf{p}(x_{1:N}, s^*_{1:N}|\theta) \leq \mathsf{p}(x_{1:N}, s^*_{1:N}|\theta^*).$$

A similar idea for estimation of parameters in a generalized linear model can be found in a recent paper by Yen [134]. The re-estimation equations can be understood intuitively — the new estimate of σ^2 is just the sample variance of the residual signal after subtracting the current piecewise linear fit from the raw data; the new estimate of p is the relative frequency of occurrence of samples where slope remains unchanged in the current fit. A formal proof is presented in Section C.2.

This result suggests an alternating maximization algorithm [135] that iterates between the slope sequence and the unknown parameters. It guarantees that the complete data likelihood increases at each step through this iterative procedure. Note that this is different from the analysis of the EM algorithm [132, 129] where the observed data likelihood function values are non-decreasing. The alternating maximization steps can now be iterated until some termination criterion is met. In most practical examples this method provides a reasonable fit in a few iterations. The complete algorithm called MAPSlope is shown in Fig. B.5 in the appendix.

Assuming there is a lower bound on the noise variance, the following result relevant to the convergence of MAPSlope algorithm can be proved:

Corollary 4.3 In addition to the hypotheses of Theorem 1, suppose there exists $\delta > 0$ such that $\sigma \geq \delta$. Then the sequence of likelihood function values obtained via alternating maximization iterations of (4.19) and (4.20) have a limit point.

A proof is presented in Section C.2. In order to make stronger claims about the properties of this limit point, further assumptions are required [136] that do not hold in this present scenario.

The re-estimation method may be applied when the smooth optimization method is used instead of MAPSlope, in which case the complete algorithm becomes a special case of alternating gradient ascent. However, the aforementioned convergence result does

not apply because the slope sequence obtained using the smooth optimization method is only an approximate maximizer.

4.5.3 Mean Squared Error Optimality Analysis and Model Order Selection

The aim of this section is to propose an empirical method for selecting the right number of quantization levels M for the set of slope values. Simple arguments from source coding theory are used to obtain upper and lower bounds for the MSE performance of the algorithm. This method can be applied directly to the sequence of estimated slope values.

Assume that the slope values originate from a continuous amplitude Markov source with amplitude levels in the interval $[0, 1]$. The goal is to characterize the performance of a decoder that outputs discrete values from the set $\{0, \frac{1}{(M-1)}, \ldots, \frac{M-2}{M-1}, 1\}$ such that the MSE is minimized. With a slight abuse of notation, let S_n denote the true slope value at sample n and let $\widehat{S_n}$ be the slope estimated from noisy data. The MSE performance metric is defined as:

$$\mathsf{MSE}(M) = \frac{1}{N} \sum_{n=1}^{N} (\widehat{S_n} - S_n)^2 \qquad [4.21]$$

where the dependence on M is due to the fact that the algorithm that generates the sequence $\widehat{S_n}$ depends on M. Other metrics such as difference between the actual and detected number of change points, and distance from the actual change points can be used in practice, but are harder to analyze theoretically.

Lower bound for MSE performance Consider an omniscient decoder (oracle) that knows the exact slope values output by the source in advance. The MSE of this decoder can be used to obtain a lower bound on the MSE obtained using any decoder. The

average error of this omniscient decoder (due to quantization) is given by [137]:

$$\begin{aligned}
\varepsilon_{LB}(M) &= \int_{0}^{\frac{1}{2(M-1)}} x^2 dx + \sum_{n=1}^{M-1} \int_{\frac{2n-1}{2(M-1)}}^{\frac{2n+1}{2(M-1)}} \left(x - \frac{n}{M-1}\right)^2 dx + \int_{1-\frac{1}{2(M-1)}}^{1} (x-1)^2 dx \\
&= \frac{1}{24(M-1)^3} + \sum_{n=1}^{M-1} \frac{1}{12(M-1)^3} + \frac{1}{24(M-1)^3} \\
&= \frac{1}{12(M-1)^2}.
\end{aligned}$$

Upper bound for MSE performance Consider an ignorant decoder that tries to minimize the MSE after discarding all input information. The MSE of this decoder provides an upper bound to the MSE performance. More "intelligent" decoders that do use the input information when deciding a slope value would have a lower MSE than this ignorant decoder. A natural strategy for the ignorant decoder is to assume that each input slope value is uniformly and randomly distributed over $[0, 1]$, and hence the best it can do is to announce a slope value that is nearest to $\frac{1}{2}$. When M is odd, the decoder has a bin at $\frac{1}{2}$ giving an MSE of $\frac{1}{12}$. If M is even, the nearest bin is always a distance of $\frac{1}{2(M-1)}$ away from $\frac{1}{2}$. So, the upper bound on the average error can be written as:

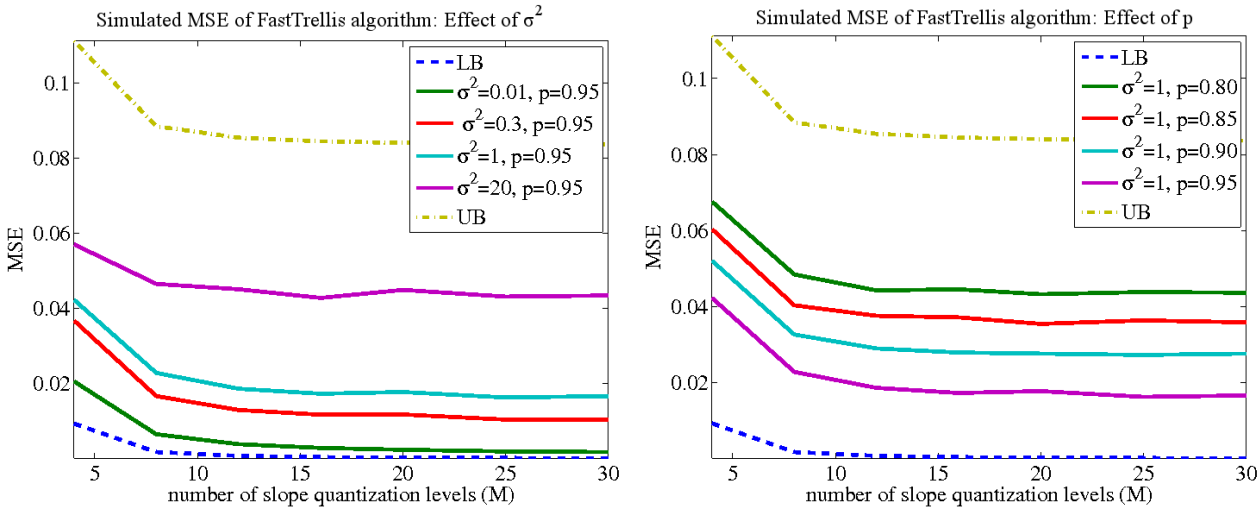
$$\begin{aligned}
\varepsilon_{UB}(M) &= \int_{0}^{1} \left(x - \frac{1}{2} - \frac{1}{2(M-1)}\right)^2 dx \\
&= \frac{1}{12} + \frac{1}{4(M-1)^2} \text{ when } M \text{ is odd,}
\end{aligned}$$

and,

$$\varepsilon_{UB}(M) = \frac{1}{12} \text{ when } M \text{ is even.}$$

Fig. 4.4(a) and 4.4(b) show the MSE through simulation on 1000 randomly generated piecewise linear datasets each of length $N = 50$. The lower and upper bounds (LB and

UB) derived above are also plotted. As one would expect, larger values of MSE are obtained for larger σ^2 values. Lower MSE values are obtained with larger values of p because the slope changes less often when p is closer to 1. Note that the true parameter values were provided as input to the FastTrellis routine during simulation.



(a) Varying noise variance

(b) Varying jump probability

Figure 4.4: MSE of the FastTrellis algorithm on randomly generated simulated data with (a) different noise variances and (b) slope change probabilities.

Two empirical conclusions can be drawn from these simulation results. First, a value of M around 15–20 is sufficient to achieve the "flat regions" of the MSE(M) curves over a reasonable range of p and σ^2 values. Using values of M larger than 20 does not give noticeable improvement in MSE. Secondly, it shows that the FastTrellis algorithm performs quite well when the underlying data is generated with p close to 1 and $\sigma^2 \leq 1$.

Other model order selection methods may be employed to choose M. The fit can be produced for a range of different values of M and the one that provides the smallest residual sum of squares can be selected. A classical parameter selection method such as leave-out-one cross-validation can also be applied. Use of information criteria such as

AIC/BIC will require modification in the setup because the M slope values are already specified in the current model; they are not estimated as part of the function-fitting algorithm.

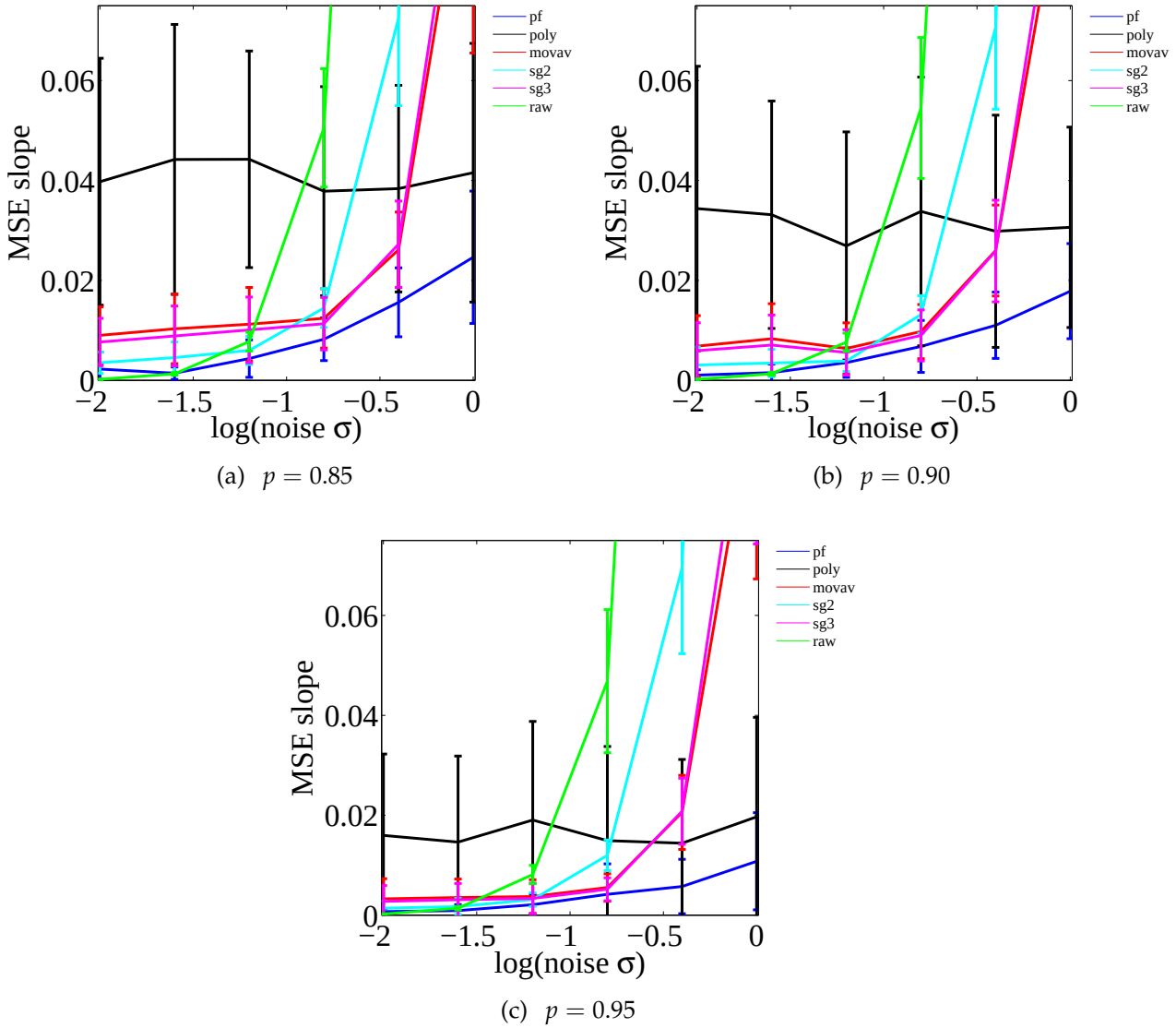
4.6 Particle Filter Results

4.6.1 Simulation Results with Synthetic Data

This section discusses simulation results involving synthetic piecewise linear data. The results shown in Fig. 4.5 show that the particle filter outperforms four other smoothing methods considered. It is not surprising that finite differencing of the raw noisy data without any smoothing has the largest MSE as seen in the three subfigures.

Note that for any constant value of p, the particle filter outperforms other filtering methods at all noise levels. It is also worth noting that the performance gap gradually decreases as p increases from 0.85 to 0.95. This is because at larger values of p there are fewer change points in the piecewise linear function (on average) and so the filtering problem is "easier" in the sense that it only has one or two straight line segments.

In practice, the parameters p and σ^2 and the upper and lower limits of the uniform distribution are not known in advance. One way to bypass problem is by estimating these values from the raw data and using these estimates as inputs to the particle filtering algorithm. The value of p is set based on the ratio of the expected number of interfaces in the physical experiment to the number of points in the dataset. The sample variance of the raw data is used as an estimate for σ^2. The lower limit for the distribution of X_n can be set to any small positive number and a reasonable upper limit can be computed from the raw data. In practice, small errors in the input values of p and σ^2 did not affect the final fit drastically. However, it is crucial that the lower limit for the uniform distribution



(a) $p = 0.85$

(b) $p = 0.90$

(c) $p = 0.95$

Figure 4.5: MSE of estimated slope values from three different noise filtering methods applied to randomly generated piecewise linear data. Simulated piecewise linear data was filtered using three different filtering algorithms (pf=particle filter, poly=4th order polynomial, movav=moving average 10 point window, sg2=Savitzky-Golay quadratic with 10 point span, sg3=Savitzky-Golay cubic with 15 point span, raw=no filtering). Local slope values were estimated by finite differencing. MSE from 50 independent simulated data vectors are presented in this figure. The particle filter provided the lowest mean MSE.

be set correctly. If this value is larger than the smallest slope in the data then the final fit would be just a straight line with a slope equal to the lower limit.

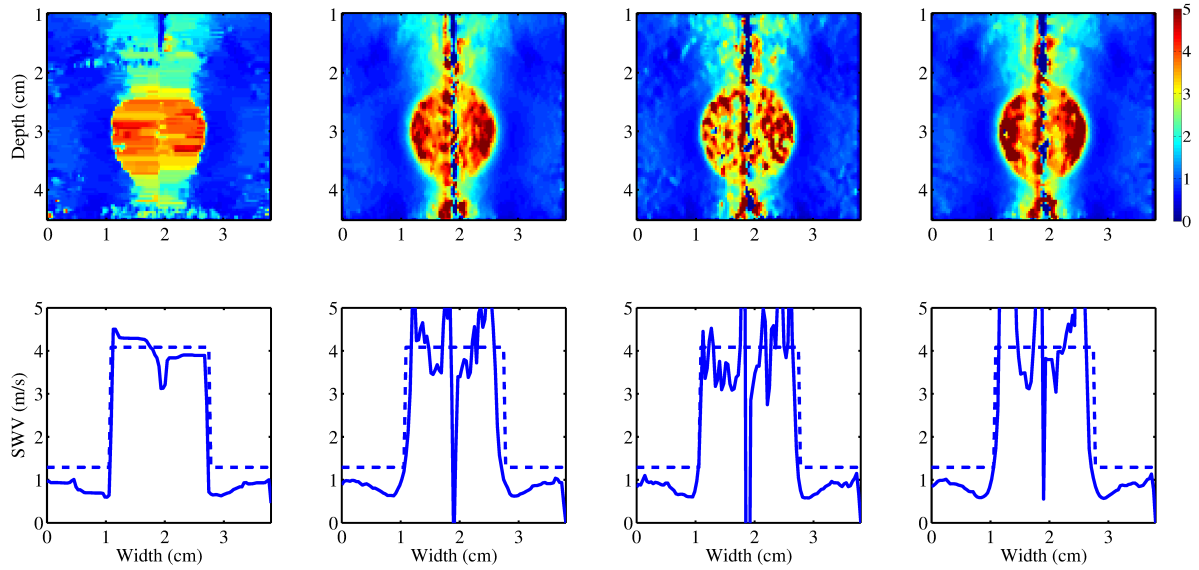


Figure 4.6: SWV maps reconstructed from data obtained from the finite element simulation model are shown here. The top row shows images reconstructed using a particle filter, moving average 10 point window, Savitzky-Golay quadratic with 15 point span, and Savitzky-Golay cubic with 20 point span, respectively from left to right. The bottom row shows SWV values along a horizontal line at a constant depth of 3 cm. True SWV profiles from the finite element model are shown with dotted lines.

4.6.2 Results from Finite Element Simulation Data

One image plane of a stiff ellipsoidal inclusion embedded in a soft background material was simulated using a finite element model in ANSYS. This is similar to the schematic shown in Fig. 3.1 without the irregularly shaped region. The shear modulus of the inclusion was set at 50 kPa whereas that of the background was set at 5 kPa. Both materials were modeled to be perfectly linearly elastic and incompressible with no

viscosity. The needle was modeled using a stainless steel rod firmly bonded to the ellipsoid, but free to slide along the surface adjoining the softer background material. More details about the model can be found in the paper by DeWall *et al.* [7].

SWV images reconstructed from the finite element simulation data are shown in Fig. 4.6. Compared to other smoothing methods, the particle filter not only preserves the sharp boundary details but also provides a lower mean squared reconstruction error as seen from the representative SWV profiles shown along a line at a constant depth of 3 cm. The spike near the center column is an artifact due to the needle.

MSE values were obtained along lines of constant depth on both sides of the needle. The box plot in Fig. 4.7 shows a summary of these values for four different methods used for obtaining SWV images from the noisy TTP image. The SWV image obtained from the noiseless TTP image was used as the ground truth for obtaining each MSE measurement. It can be seen that the particle filter provides the lowest median MSE with smallest spread among the methods considered.

Variance of the estimated SWV values calculated using Eq. (4.11) SWV values can be displayed for the clinician alongside the SWV image as seen in Fig. 4.8. Note the stray high variance pixels occur at locations laterally away from the needle where the shear wave pulse is more difficult to track due decreased peak displacement. However, most of the pixels have a variance close to zero indicating that the SWV estimate are reliable.

4.6.3 Experimental Results from TM Phantoms

Ten independent datasets were obtained from each of the two TM phantoms using the experimental setup described previously in Section 3.2. In order to gauge the performance of the particle filtering algorithm vis-a-vis existing imaging and data processing techniques, SWV estimates were obtained from two other methods. In the first

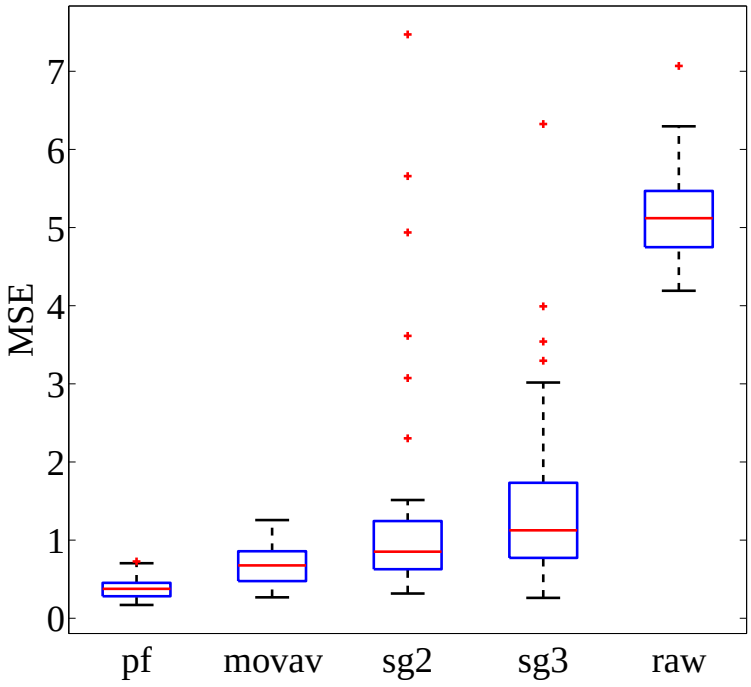
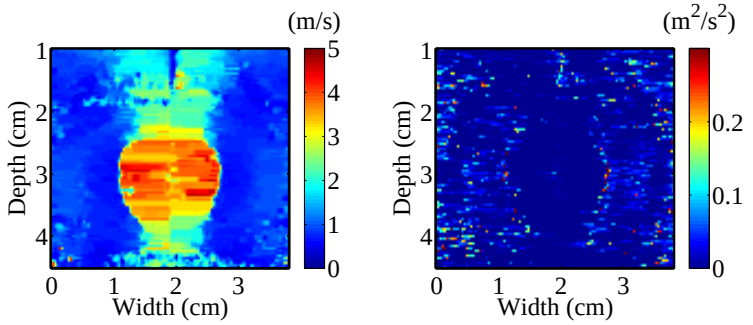


Figure 4.7: A finite element simulation model was used to export frame-to-frame displacements which were processed using various algorithms to estimate SWV. Mean squared reconstruction error in the SWV maps produced from the finite element simulation model are shown here, where the SWV from the finite element model was used as ground truth. (pf=particle filter, movav=moving average 10 point window, sg2=Savitzky-Golay quadratic with 15 point span, sg3=Savitzky-Golay cubic with 20 point span, raw=no filtering)

method, a simple least-squares fourth order polynomial fit was used to filter the raw TTP data. Young's modulus was calculated using (1.1). SWV and shear modulus estimates obtained from these four different methods (EVE with particle filtering, EVE with least-squares filtering, SSI and mechanical testing) are summarized in Table 4.1. Test metrics were calculated for the SWV maps obtained using both the particle filter and the least-squares filter. These are shown in Table 4.2. Standard deviations are calculated from individual SNR, C and CNR values in dB obtained from each of the ten datasets.



(a) SWV (b) SWV variance

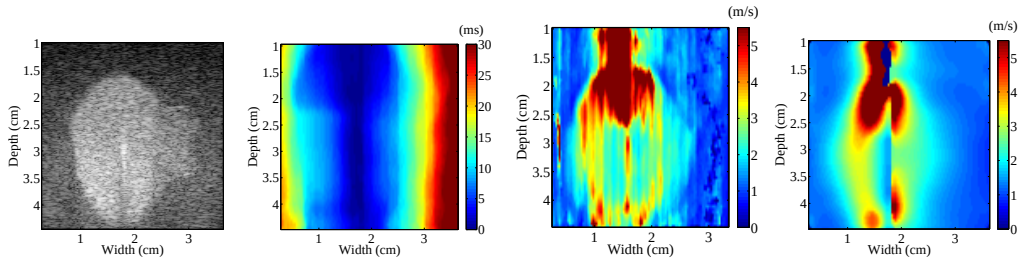
Figure 4.8: Since the SWV values are obtained using a Bayesian model, the posterior density can be used to produce a variance image that provides feedback about the reliability of the SWV estimates. The variance image in (b) is calculated using Eq. (4.11) for each pixel in the SWV image in (a). The SWV image is generated using finite element data.

Table 4.1: Shear wave velocity and Young's modulus estimates using particle filter

		Phantom-1			Phantom-2		
		e	t	b	e	t	b
EVE	SWV PF (m/s)	3.07 ± 0.7	2.02 ± 0.32	0.91 ± 0.31	4.68 ± 1.3	2.99 ± 0.4	1.32 ± 0.68
	SWV LS(m/s)	3.14 ± 0.79	1.96 ± 0.23	1.16 ± 0.12	4.56 ± 0.75	3.1 ± 0.32	2.09 ± 0.2
	E PF (kPa)	29.6 ± 13	12.6 ± 3.8	2.74 ± 1.5	70.6 ± 52	27.3 ± 6.9	6.62 ± 9.5
	E LS (kPa)	31.5 ± 18	11.7 ± 2.7	4.07 ± 0.86	63.9 ± 21	29.2 ± 6	13.2 ± 2.4
SSI	SWV (m/s)	2.8 ± 1.1	2.3 ± 0.8	1.3 ± 0.4	4.1 ± 1.8	2.4 ± 1.4	2.1 ± 0.7
	E (kPa)	24.2 ± 5.8	13.3 ± 3.5	4.8 ± 0.5	50.1 ± 10.5	17.6 ± 4.8	13.6 ± 1.6
ELF	E (kPa)	56.57 ± 0.25	24.74 ± 0.63	4.55 ± 0.06	41.05 ± 0.20	20.52 ± 0.54	15.10 ± 0.80

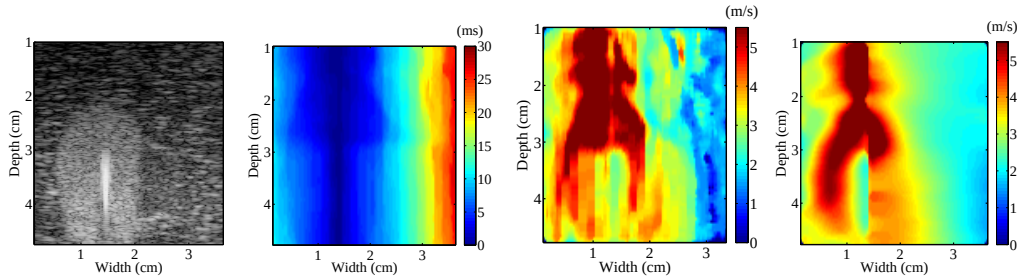
Values of SWV and Young's modulus of different regions in the two phantoms obtained from particle filtering (PF) and least-squares filtering (LS) applied to the TTP data. Corresponding values obtained from the Supersonic Imagine scanner and mechanical testing are also indicated. (e = ellipsoidal inclusion, t = irregular tumor region, b = background).

Sample outlines are shown in Figs. 4.11 and 4.12. Area estimate results are summarized in Table 4.3.



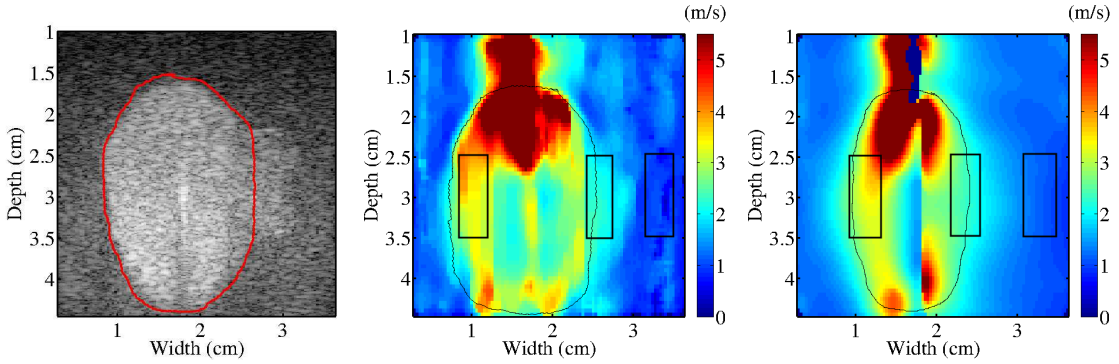
(a) B-mode (b) TTP (c) SWV PF (d) SWV LS

Figure 4.9: Results from shear wave tracking in Phantom-1 are shown. (a) B-mode image of the phantom, (b) TTP map for the shear wave pulse at various locations in the phantom, (c) SWV map generated by applying the particle filtering (PF) algorithm to denoise the TTP data followed by lateral gradients to obtain wave velocity, (d) same as (c) but with a least-squares (LSQ) polynomial fit instead of the PF.



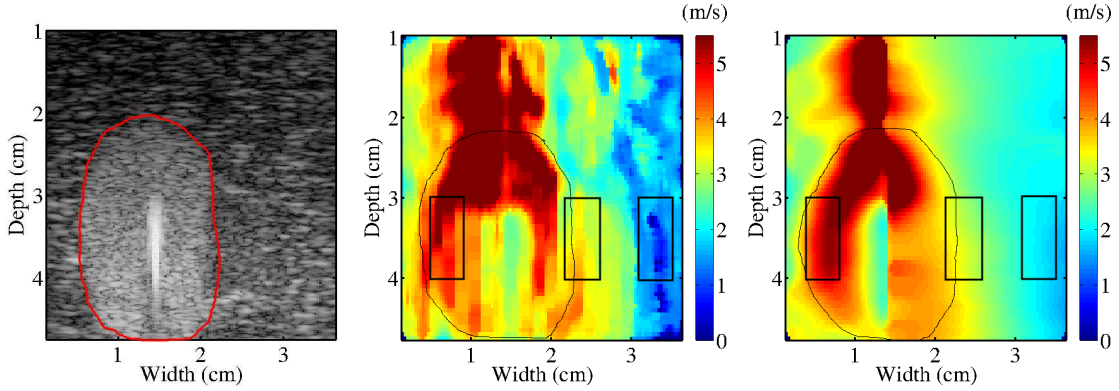
(a) B-mode (b) TTP (c) SWV PF (d) SWV LSQ

Figure 4.10: Results from shear wave tracking in Phantom-2 are shown. (a) B-mode image of the phantom, (b) TTP map for the shear wave pulse at various locations in the phantom, (c) SWV map generated by applying the particle filtering (PF) algorithm to denoise the TTP data followed by lateral gradients to obtain wave velocity, (d) same as (c) but with a least-squares (LSQ) polynomial fit instead of the PF.



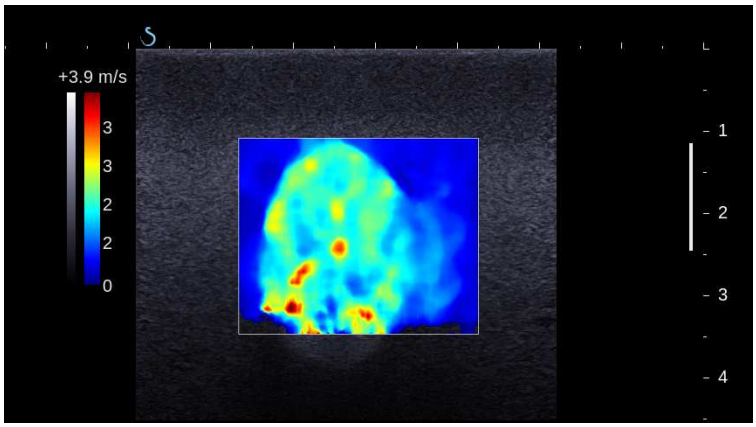
(a) B-mode (b) SWV PF (c) SWV LSQ

Figure 4.11: ROIs and inclusion boundaries used for Phantom-1 are shown. Boundary used for B-mode area estimation is shown in (a). Boundaries for area estimation and ROIs used for calculating the various statistics on the SWV maps are shown in (b) and (c).

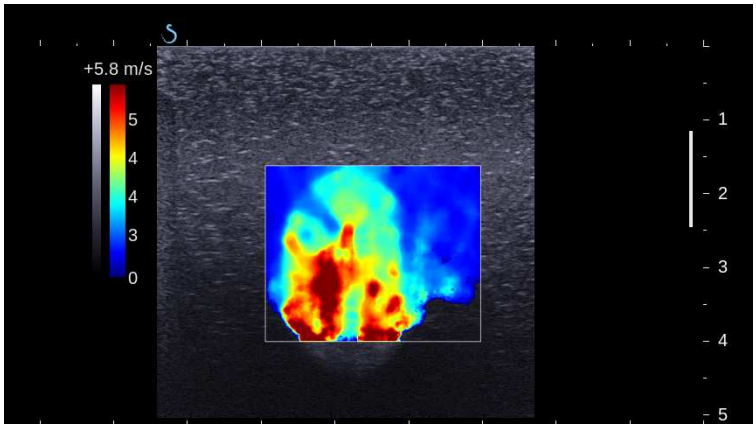


(a) B-mode (b) SWV PF (c) SWV LSQ

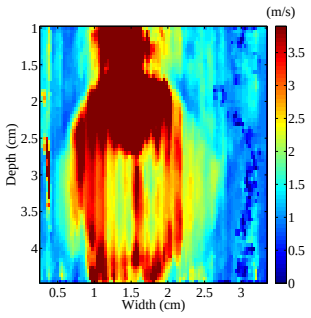
Figure 4.12: ROIs and inclusion boundaries used for Phantom-2 are shown. Boundary used for B-mode area estimation is shown in (a). Boundaries for area estimation and ROIs used for calculating various statistics on the SWV maps are shown in (b) and (c).



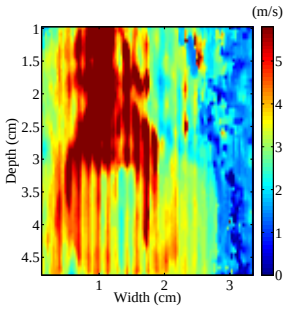
(a) SWV map of Phantom-1 using SSI



(b) SWV map of Phantom-2 using SSI



(c) SWV map of Phantom-1 using particle filter



(d) SWV map of Phantom-2 using particle filter

Figure 4.13: SWV maps obtained using the clinical software interface of the Supersonic Imagine Aixplorer scanner using SSI are shown. Results from Phantom-1 and Phantom-2 are shown in (a) and (b), respectively. Reconstructions using the particle filtering algorithm are shown again on the same SWV scale for comparison in (c) and (d).

Table 4.2: SNR, C and CNR particle filter

	Regions	Method	Phantom-1	Phantom 2
SNR (dB)	e	PF	13.2 ± 1.2	13.2 ± 3
		LS	12.4 ± 1.3	16.8 ± 2.7
	t	PF	16.6 ± 2	18.4 ± 2.5
		LS	19.8 ± 1.1	20.3 ± 1.2
	b	PF	9.67 ± 1.6	6 ± 1.8
		LS	20.6 ± 1.8	23.4 ± 1.7
C (dB)	e/t	PF	3.62 ± 0.56	3.83 ± 0.67
		LS	4.1 ± 0.6	3.39 ± 0.49
	t/b	PF	7.06 ± 0.82	7.15 ± 0.77
		LS	4.56 ± 0.63	3.5 ± 0.53
	e/b	PF	10.7 ± 0.7	11 ± 1
		LS	8.66 ± 0.6	6.89 ± 0.85
CNR (dB)	e/t	PF	11.8 ± 2.8	12.8 ± 3.7
		LS	13.1 ± 2.7	17.9 ± 4.2
	t/b	PF	22.9 ± 3.6	19.7 ± 2.7
		LS	27.5 ± 1.8	25.5 ± 1.9
	e/b	PF	24.6 ± 1.8	23 ± 4.1
		LS	22.6 ± 2.2	28.5 ± 5.1

Values of SNR, C and CNR obtained for various pairs of regions in the two phantoms. All numbers are in dB. (e = ellipsoidal inclusion, t = irregular tumor region, b = background).

Results of EVE and B-mode scans of the two phantoms are shown in Figs. 4.9 and 4.10. Observe that in the SWV reconstruction the irregular tumor area can be distinguished from the stiff inclusion and the background material. For comparison, the SWV image generated using a simple least-squares fourth order polynomial fitting method is also shown. Although there is greater noise reduction and smoothing in the least-squares fit, the boundary details of the inclusion and partially ablated region get smeared out. For comparison, SWV maps obtained using the SSI technique are also shown in Fig. 4.13.

4.7 Piecewise Linear Fitting Results

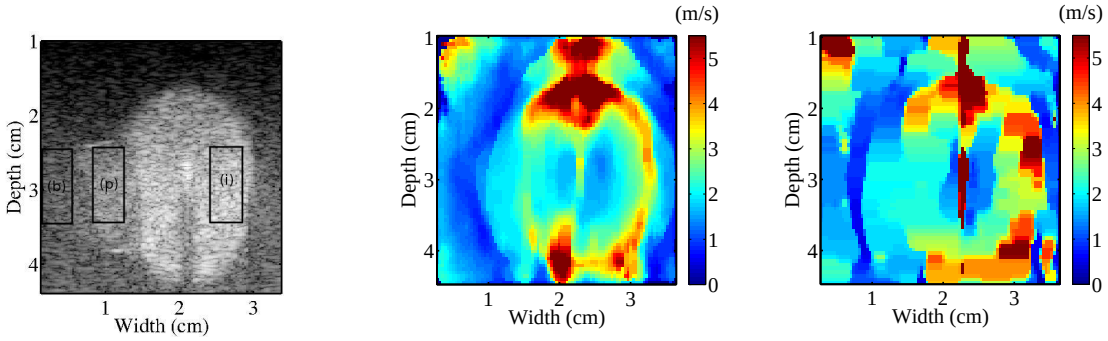
A representative SWV map reconstructed using each of the two algorithms along with a B-mode ultrasound image of the phantom is shown in Fig. 4.14. The two algorithms

Table 4.3: Inclusion area estimates particle filter method

Method	Phantom-1	Phantom-2
SWV PF	4.45 ± 0.15	4.13 ± 0.18
SWV LS	4.11 ± 0.19	4.01 ± 0.14
B-Mode	4.68 ± 0.14	4.47 ± 0.11

Estimates of area in cm^2 of the ellipsoidal inclusion obtained from the SWV maps and B-mode images are shown. Areas were estimated by manually outlining the inclusion and calculating the mean and standard deviation of the areas over 10 independent datasets.

perform equally well in terms of visual delineation of the outer inclusion boundary. However, the reconstruction obtained using piecewise-linear regression appears blotchy as compared to the particle filtered image. This is to be expected because the particle filter has the freedom to choose slope values smoothly in a specified interval, whereas the piecewise-linear regression algorithm is forced to approximate the fit with a few straight lines with abrupt breakpoints.



(a) B-mode (b) Particle filter (c) Piecewise-linear filter

Figure 4.14: B-mode image of the phantom consisting of inclusion (i), partially ablated region (p) and background (b) is shown in (a). SWV maps reconstructed using a (b) particle filter algorithm and (c) piecewise-linear fitting algorithm are also shown.

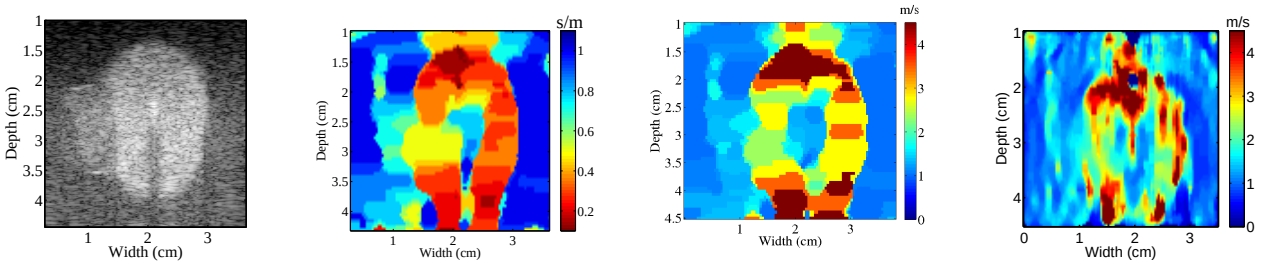
ROIs of size 10 mm $\times$ 5 mm are fixed in each of the three regions of the phantom to obtain quantitative estimates of SWV and Young's modulus. The locations of these ROIs are shown overlaid on the B-mode image in Fig. 4.14(a). The means and standard deviations averaged over the ROIs and averaged over ten independent data sets are shown in Table 4.4. The mean SWV estimates obtained from the two methods differ by less than 10% in all the three stiffness regions of the phantom. The mean Young's modulus estimates of the background are within 1 kPa of those obtained from the ELF mechanical testing system, and 15 kPa for the inclusion.

Table 4.4: Comparison of SWV estimates particle filter vs. piecewise linear filter

ROI	SWV (m/s)		E (kPa)		
	PF	PL	PF	PL	ELF
Inclusion	3.8 ± 2.2	3.4 ± 1.5	57.2 ± 70	42.2 ± 58	54.4 ± 0.1
Partially ablated	2.0 ± 0.2	2.0 ± 0.3	11.9 ± 2.6	12.1 ± 4.2	21.6 ± 0.3
Background	1.3 ± 0.2	1.4 ± 0.4	5 ± 1.9	6.5 ± 6.1	4.8 ± 0.5

Values of SWV and Young's modulus of different regions in the two phantoms obtained from particle filtering (PF) and piecewise-linear filtering (PL). Stiffness values obtained from mechanical testing (ELF) are also indicated.

Certain limitations arise from the simplified shear wave propagation model. Viscosity and frequency dispersion can cause the wave pulse shape to change. Both SWV maps show low velocity artifacts around the needle inside the ellipsoidal inclusion. This is probably because the wave pulse takes a certain finite time duration to attain its peak velocity after the needle is vibrated. High velocity artifacts are seen above and below the inclusion because of the assumption that the shear wavefront travels purely in the lateral direction. Also observe in Table 4.4 that the variance of the velocity and Young's modulus estimates are quite high in the stiff region of the phantom. This is because the ATP corresponding to the stiffer region has a smaller slope value that is difficult to estimate accurately [70, Eq. (21)].



(a) B-mode (b) Slowness (c) Velocity (d) Velocity (LOWESS)

Figure 4.15: Results from shear wave tracking using the MAPSlope algorithm. (a) Ultrasound B-mode image of the phantom, (b) slowness map generated by applying the MAPSlope algorithm to denoise the arrival time data and obtain MAP slope estimates. Wave velocities are shown in (c) by calculating the reciprocal of the slowness values. The velocity values estimated using a 2nd order LOWESS filter with a 15 sample sliding window is are shown in (d).

4.8 MAPSlope-FastTrellis Results

Results of the complete procedure are shown alongside a B-mode ultrasound image in Fig. 4.15. Note that the three regions of different stiffnesses are visible in the B-mode scan in Fig. 4.15(a); this is done on purpose by altering the acoustic echogenicity of the TM material used. As mentioned previously, these stiffness variations are not easily visible in B-mode scans of real tissue. The slowness map of a shear wave pulse tracked in the same imaging plane is shown in Fig. 4.15(b). There is good correlation between the boundary of the stiff ellipsoidal region The small irregular area on the left of the ellipsoid which has an intermediate stiffness is also visible in the slowness map. Similar visualization is possible through the shear wave velocity map in Fig. 4.15(c). Additionally, a LOWESS [72] filtered image is shown in Fig. 4.15(d). It uses a local quadratic model with a sliding window of 15 samples. This kind of "least-squares smoothed" slope estimation is common in shear wave imaging literature [7]. It can be seen that the boundary details are sharper when the MAPSlope algorithm is used.

Numerical results obtained with three regions of interest (ROI), each of size 1 cm $\times$ 2 cm, fixed in the slowness and shear wave velocity maps are shown in Table 4.5. Standard image quality assessment metrics from the ultrasound elastography literature are used for this study [96]; Table 4.6 shows these evaluation metrics. For each region, the signal to noise ratio (SNR) is defined as:

$$SNR = \frac{\mu}{\sigma}$$

where μ and σ respectively denote the mean and the standard deviation values calculated over the ROI. The contrast (C) between a pair of regions is defined as:

$$C = \frac{\mu_1}{\mu_2}$$

Table 4.5: Slowness and stiffness estimates MAPSlope-FastTrellis algorithm

	Stiff	Intermediate	Soft
Slowness (s/m)	0.356 ± 0.1	0.537 ± 0.11	0.859 ± 0.18
Velocity (m/s)	3.09 ± 0.9	2.03 ± 0.35	1.21 ± 0.29
SNR (velocity)	12.6 ± 3.7	17.6 ± 3.4	12.7 ± 1.9
E (kPa)	30.4 ± 22	11.5 ± 3.8	4.86 ± 2.9

Values of shear wave slowness, shear wave velocity and Young's modulus for the three different regions in the experimental phantom obtained from the MAPSlope algorithm are indicated.

where the subscripts indicate the two different ROIs. Similarly, the contrast to noise ratio (CNR) is defined as [138]:

$$CNR = \frac{2(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}.$$

Observe from Table 4.5 that the SNR values calculated from the shear wave velocity maps are around 10 dB in all three regions of the phantom, which is quite high given the noise levels of raw ultrasound echo data. It is worth noting in Table 4.6 that the best contrast and contrast-to-noise ratios are obtained for the pair of regions that differ the most in shear stiffness.

4.9 Discussion

Observe from Fig. 4.9 that the particle filtering method applies an optimal amount of smoothing to the raw TTP data and produces good boundary delineation even between the inclusion and the irregular tumor regions that do not differ greatly in their shear modulus values. There is always a risk of over or undersmoothing when using *ad hoc* function fitting algorithms (like least-squares) that may blur boundary details. Quantitative estimates of SWVs and Young's moduli obtained using the particle filter agree well with the ground truth obtained from mechanical testing and least squares filtering.

Table 4.6: Image quality metrics MAPSlope-FastTrellis algorithm

	Stiff/Inter.	Inter./Soft	Stiff/Soft
C	3.6 ± 0.77	4.57 ± 0.72	8.17 ± 0.45
CNR	10.3 ± 6.4	18.1 ± 4	21.1 ± 6.6

Contrast (C), signal-to-noise ratios (SNR) and contrast-to-noise ratios (CNR) (in dB) obtained from shear wave velocity estimates for three pairs of regions are shown. See text for definitions of these quantities. The standard deviations are calculated from the dB values obtained over each ROI from individual datasets.

The SNR, C and CNR values indicate that the particle filter is within a few dB of the least-squares technique in suppressing noise, while providing better visual delineation between stiffer and softer areas in the phantom. The particle filtered SWV maps have quite high SNR and CNR values of at least 30 dB and 40 dB respectively. These test metrics do not account for any measurement bias that may be present in the raw TTP data and are used only to compare the performance of the two filtering methods.

Finally, inclusion area estimates obtained from the particle filtered SWV maps are quite close to those obtained from B-mode imaging. As seen from Table 4.3, the least-squares method underestimates the inclusion size which may be a side effect of over-smoothing. It is worth noting that the contrast between the three different regions of the phantom easily visible in B-mode is because of intentionally increased backscatter contrast in these manufactured phantom materials. In real tissue, differentiating stiffer and softer areas using B-mode scans is often challenging due to mixed echogenic contrast [139]. Area estimates may be susceptible to user variability because of the manual outlining step. Therefore, these measurements should not be used as the primary metric for comparing the performance of these SWV image reconstruction algorithms.

Young's modulus estimates of the stiffer areas in Phantom-1 that are obtained from mechanical testing do not agree with the results from ultrasound elastography estimates

obtained from EVE and SSI. There is much better agreement of modulus estimates for the data obtained from Phantom-2. This may be because Phantom-1 is over a year old and there is a possibility of gradual degradation of the stiffness of phantom materials over time [140]. Cylindrical samples that used in mechanical testing are stored separately and hence are not under identical physical conditions as the material in the actual phantom. Moreover, the numbers obtained from elastography may have an inherent bias because the raw data undergoes multiple smoothing and filtering operations before obtaining these modulus estimates.

SWV maps obtained using the Supersonic Imagine scanner are shown in Fig. 4.13. It is apparent from these images that EVE has the ability of generating SWV maps at greater depths and larger fields of view. Due to imaging depth limitations, inclusion area estimates could not be obtained using SSI. The proprietary velocity reconstruction algorithm appears to use a greater degree of smoothing. SWV estimates appear lower than those obtained with the particle filter. Nevertheless, the SSI technique differentiates various regions in the phantoms quite well, especially in case of Phantom-1 as seen in Fig. 4.13(a).

The ability of EVE technique to resolve fine boundary details is limited by the shear wave pulse width used. The data shown here was generated using a 30 ms wide shear wave pulse. Although it may be possible to obtain sharper delineation by using a narrower pulse, it was observed during experimentation that due to certain physical limitations of the actuator system, accurate amplitude control could not be obtained for shorter vibration durations. The effect of the width of the pulse on estimation of TTP values and subsequent effect on the resolution of the SWV images will be analyzed in the future.

The data processing algorithms used throughout this work assume that there is pure lateral propagation of the wave throughout the phantom. This assumption fails to hold for regions above and below the inclusion because the needle is bonded only to the interior of the inclusion [141]. As a result, SWV artifacts can be seen in the regions that are shallower or deeper than than the stiffer ellipsoid. In order to focus attention on the data obtained from the regions that are laterally adjacent to area where the needle is bound to the phantom material, the SWV maps in Fig. 4.9 and Fig. 4.10 are shown beginning at a depth of about 1 cm. The wave propagation phenomenon above and below the stiff ellipsoid is more complex than pure transverse wave motion and a separate study to analyze this aspect may be necessary.

The data acquisition system is capable of operating at a frame rate of about 2 kHz which provides sufficient time sampling to track the shear wave pulse. As a practical matter, the particle filter runs slower than least-squares polynomial fitting. However, the algorithm is parallelizable because data at each depth is filtered independently. The current test implementation takes a few minutes to process the images shown in Figs. 4.9 and 4.10. It is quite common to see a speedup by an order of magnitude when implemented as compiled code, providing almost real-time monitoring capability.

5

Three Dimensional Reconstruction with Sheaf Acquisitions

> He that goeth forth and weepeth,
> bearing precious seed, shall doubtless
> come again with rejoicing, bringing
> his sheaves with him.
>
> Psalm 126:6

5.1 Introduction

In this chapter, two dimensional (2D) electrode vibration elastography (EVE) is extended to three dimensions (3D) by utilizing radiofrequency (RF) echo signals acquired over a "sheaf" of imaging planes. A sheaf is defined as a collection of planes that intersect along a common axis. The 3D reconstruction algorithm is termed "**S**heaf **O**f **U**ltrasound **P**lanes **R**econstruction" or **SOUPR**. This cylindrically symmetric method of acquisition is naturally suited to electrode vibration elastography where shear wavefronts travel outward from a vibrating needle which acts as a line source of shear waves [6]. Moreover, the overall shape of the ablation is approximately cylindrically symmetric with the ablation needle coinciding with the axis of symmetry.

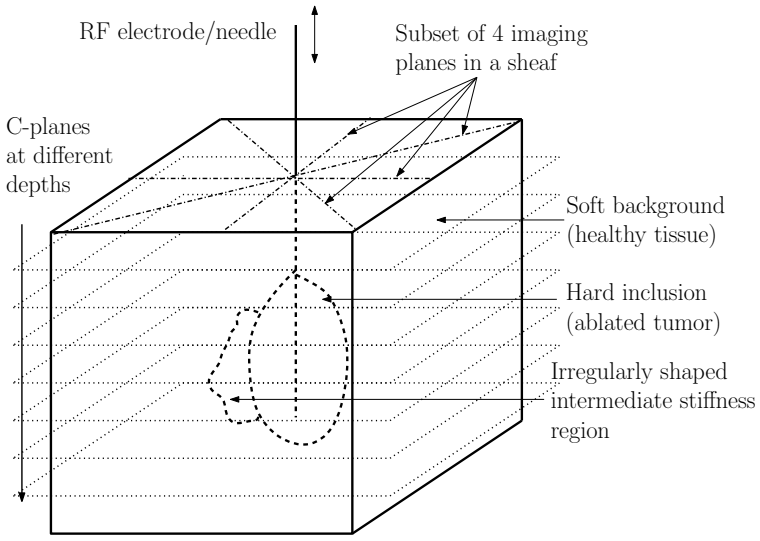


Figure 5.1: A subset of a sheaf of imaging planes containing four equiangular planes passing through a line that coincides with the ablation needle is shown. A stack of C-planes over which the shear wave velocities are interpolated is also shown. The needle is vibrated vertically to set up a shear wave pulse that travels outwards and away from the needle in a cylindrically symmetrical fashion.

Ideally, a full 3D reconstruction algorithm must induce and track shear wave displacements in 3D by acquiring full volume RF ultrasound echo data. However this requires the use of matrix array transducers [142, 12] which are more expensive than traditional 1D linear and phased array probes and may not provide high frame rates to track shear waves. Sheaf acquisition can be used to bypass these limitations by generating SWV image slices using data from linear or phased array transducers and assembling them into a 3D visualization.

This chapter presents details of the data acquisition setup used for experimental validation of SOUPR. In order to reconstruct a 3D shear wave velocity (SWV) reconstructions, data is acquired over a subset of a sheaf of planes. This subset can be chosen in such a way that the planes are equispaced in angle. For example, a sheaf of four imaging planes is shown in Fig. 5.1. SOUPR can also handle scattered data points from irregularly spaced imaging planes. The transducer is manually adjusted to image specific

angular locations in the sheaf. Guide markers on the tissue-mimicking (TM) phantom container walls are used to align the angle of the imaging plane. The effect of the number of imaging planes on reconstruction quality is studied using standard image quality metrics. This method relies on misregistration errors being small enough to be considered as part of the additive measurement noise. In general, the method may fail if there is significant relative translation or rotation between two volumes. If greater accuracy in transducer placement is desired, feedback control systems like one described by Abolmaesumi *et al.* [143] may be employed. Such systems are beyond the scope of this work.

5.2 Reconstruction Algorithm

The 3D reconstruction algorithm SOUPR consists of four distinct data processing steps, starting with the acquisition of beamformed RF ultrasound echo data, finally leading to the 3D stack of C-planes of SWV estimates:

1. Displacement estimation from beamformed echo data over each imaging plane.
2. Wavefront localization to obtain time to peak (TTP) [54] displacement at different locations in the imaging plane.
3. SWV estimation from TTP plots to obtain SWV maps over each imaging plane.
4. Smooth function approximation on a grid over each C-plane.

As seen in Fig. 5.1, each C-plane is perpendicular to all the imaging planes. The SOUPR algorithm presented here effectively breaks down a 3D reconstruction problem into a sequence of decoupled 2D function approximation problems. Each C-plane is processed separately and the final 3D reconstruction is generated by stacking them together.

5.2.1 Displacement Estimation

Ultrasound echo data is acquired over each imaging plane as described previously in Section 3.3.1. The sequential tracking acquisition enables axial strips to be assembled into individual frames to obtain high frame rates for tracking shear waves. These frames provide snapshots of the underlying medium at different time instants. Frame-to-frame displacements are estimated using a standard 1D cross-correlation algorithm [144]. An axial displacement estimation routine is used with a window length of 2 mm and 75% overlap. This procedure is repeated over the entire imaging plane to obtain a displacement vs. time profile at each pixel [145].

5.2.2 Wavefront Localization

The high frame rate displacement data is used to localize the shear wave pulse. It is assumed that the wave travels purely laterally away from the needle. The time of arrival of the wave at different locations away from the needle is recorded by finding the time of peak displacement [54, 53]. A frequency domain filter discards any frequency components that are smaller than 10% of the largest component of the frequency spectrum of the displacement vs. time profiles. This filters out any "high frequency" noise components in the displacement-time profiles making it easier to locate the peak. A quadratic fit with a 5-point window is used around the peak to get sub-frame-number resolution for the location of the peak.

5.2.3 Imaging Plane Reconstruction

The piecewise linear fitting algorithm described in Section 4.2 is used to reconstruct the shear wave velocity images on each imaging plane. The image plane containing the

partially ablated region is also imaged using a commercial Supersonic Imagine (Aix-en-Provence, France) scanner using the ShearWave^{TM} Elastography mode. SWV estimates are obtained using ROIs placed in the three different regions of the phantom.

SWV estimates on each C-plane are obtained from a second function approximation routine. Fig. 5.1 also illustrates a set of C-planes over which the SWV values are interpolated. The goal is to provide smooth reconstructions of SWV values using a fine grid on every C-plane.

Depending on the number of planes imaged in the sheaf, SWV estimates are available along many concurrent radial lines passing through the needle axis on each C-plane. (However, it is worth noting that this step of the SOUPR algorithm is quite general and can be applied even if data is not acquired in a sheaf pattern.) For convenience, let $\mathbf{x}$ denote a vector of the unknown values of SWVs on the user defined C-plane grid and let $\mathbf{b}$ denote the vectorized version of the known SWV values (data) along radial lines on the same C-plane. In general, the known data points do not coincide with grid points. Since this is now a 2D smoothing problem on a particular C-plane, a local bilinear interpolation scheme is used. Each known function value is expressed as a linear combination of the four neighboring grid points. The weighting coefficients for the grid neighbors of each data point are represented using rows of a matrix $\mathbf{A}$. The number of rows in $\mathbf{A}$ is equal to the length of $\mathbf{b}$ and the number of columns is equal to the length of $\mathbf{x}$. So, the unknown function values on the grid can be computed by solving the system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$.

5.2.4 Regularized C-plane Reconstruction

Note that the number of unknowns (grid locations) in the problem can be much larger than the number of points at which the stiffness estimates are actually known. So

$\mathbf{A}$ may not be full rank (it has fewer rows than columns) which makes the system of linear equations ill-posed. This ill-posedness is circumvented by adding an extra term which penalizes a large value of the derivative at any point in the reconstruction. Since differentiation in the case of discrete data is simply a finite difference operation, it can be compactly represented using another (square) matrix $\mathbf{B}$. Second order central differences (Laplacian) are used here to penalize the second derivative of the underlying function. The derivative at each grid node can be expressed as a linear combination of its four neighboring grid nodes. The following least-squares optimization problem can be used to solve for the unknown values on the grid:

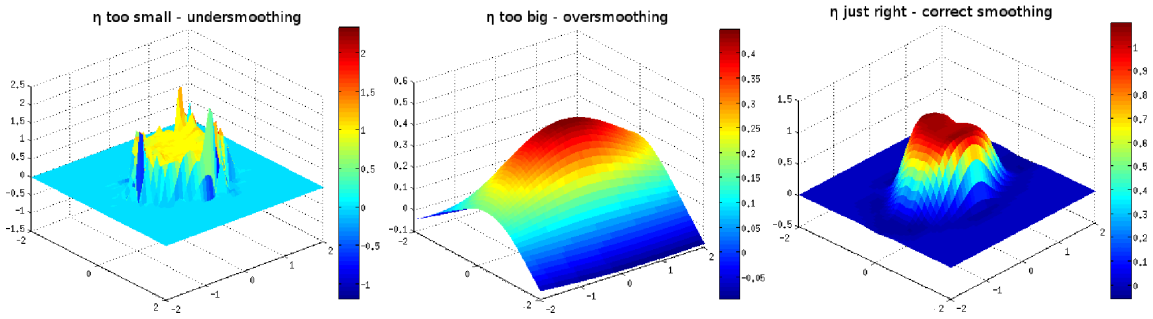
$$\arg\min_{\mathbf{x}} \; ||\mathbf{A}\mathbf{x} - \mathbf{b}||^2 + \eta||\mathbf{B}\mathbf{x}||^2 \qquad [5.1]$$

where $\eta > 0$ is a regularization parameter that controls the amount of smoothing. The closed form solution

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \eta\mathbf{B}^T\mathbf{B})^{-1}\mathbf{A}^T\mathbf{b} \qquad [5.2]$$

can be obtained by setting the derivative of the objective function in (5.1) with respect to $\mathbf{x}$ to zero and solving for $\mathbf{x}$ (see Section C.3).

The aforementioned problem setup is closely related to the idea of "Tikhonov regularization" for ill-posed problems [146, Ch. 8]. Explicit computation of the matrix inverse in (5.2) should be avoided in practice. For the present setup, each row of $\mathbf{A}$ and $\mathbf{B}$ contains at most four and five non-zero entries respectively, and the matrix in the parentheses in (5.2) is sparse, symmetric and block diagonal which can be efficiently inverted using sparse factorization algorithms [113]. This also bypasses any computational issues with large matrix dimensions thereby allowing use of fine reconstruction grids.



(a) Under-smoothing (b) Over-smoothing (c) Correct smoothing

Figure 5.2: This figure shows the effect of using different values of the smoothing parameter η in the fitting procedure applied to simulated noisy data. When η is too small, the resulting fit follows the noise as seen in (a), whereas if η is too big it causes over-smoothing as seen in (b). The correct smoothing parameter value strikes an "optimal" balance as seen in (c). This value is chosen to minimize the OCV score plotted in Fig. 5.3.

5.2.5 Smoothing Parameter Selection

If the smoothing parameter η is chosen arbitrarily, larger values will force the final fit to be closer to the null space of the penalty function which induces smoothness, whereas small values will result in a more undulating fit. However an objective method for selecting this parameter is necessary for studying reconstruction quality using summary statistics like the mean and standard deviation calculated from regions of interest in the reconstructed image. For this study, an automatic method called leave-one-out ordinary cross-validation (OCV) is used to circumvent any user induced variability from different values of η. In this method, the fitting routine is repeatedly run by excluding one data point at a time and calculating the OCV score function given by:

$$\mathsf{OCV}(\eta) = \frac{1}{M}\sum_{k=1}^{M}(\widehat{b}_k - b_k)^2$$

where M denotes the number of data points, b_k denotes the *known* data value at a skipped data point location k, and $\widehat{b}_k$ denotes the value *predicted* by the optimization

procedure when all but the k^{th} data point are used to solve (5.1). A grid search is then used to arrive at the choice of η that minimizes this score function. The smoothing parameter chosen in this manner also has a desirable statistical property of minimizing the cross-validation score which is an unbiased estimator of the mean squared error risk [147, Sec 5.3]. Example C-plane reconstructions with three different values of η are shown in Fig. 5.2. These plots were generated using a test function of two dimensions: $f(x, y) = 1$ when $x^2 + y^2 \leq 1$ (unit circle centered at the origin), and zero otherwise, with the domain $-2 \leq x, y \leq 2$. Gaussian noise was added to this function to simulate measurement errors in the shear wave velocity and registration errors due to misalignment of the image plane angle. The function values were then imputed on a grid using (5.2). The "correct" value of η for this case was found by minimizing $\mathrm{OCV}(\eta)$ shown in Fig. 5.3. In the interest of processing time, η was estimated only once per dataset using the C-plane at half the maximum depth. It can be re-estimated for every C-plane if desired.

Intuitively, the problem of choosing the sampling resolution for each imaging plane and number of planes in the sheaf depends on the "roughness" or frequency content of the true underlying signal to be reconstructed. From a signal-theoretic perspective, this can be formalized into a sampling theorem described in Section C.4. In order to better understand the effect of noise and varying number of image planes on the reconstruction quality of the SOUPR algorithm, a Monte Carlo simulation experiment was performed using a single C-plane. Using the same test function as above, synthetic data is generated with 4, 6, 12 and 16 concurrent radial lines with 80 equidistant samples along each line. Gaussian noise for different signal to noise ratios (SNR) was added to simulate measurement errors. Since the maximum test function value is 1, SNR is defined as $20 \log_{10}(1/s)$ where s is the standard deviation of simulated Gaussian noise. This noisy

data was processed with the SOUPR algorithm to reconstruct C-plane function values $\hat{f}(x,y)$ on a 100×100 grid in the domain $-2 \leq x,y \leq 2$. Reconstruction mean squared error was calculated using $\frac{1}{10^4}\sum_{i=1}^{100}\sum_{j=1}^{100}\left|f\left(\frac{4i-202}{99},\frac{4j-202}{99}\right)-\hat{f}\left(\frac{4i-202}{99},\frac{4j-202}{99}\right)\right|^2$, where the scaling and shifts ensure that the functions f and $\hat{f}$ are evaluated in the domain $-2 \leq x,y \leq 2$. This simulation was repeated 10,000 times. Results are shown in Section 5.3.1.

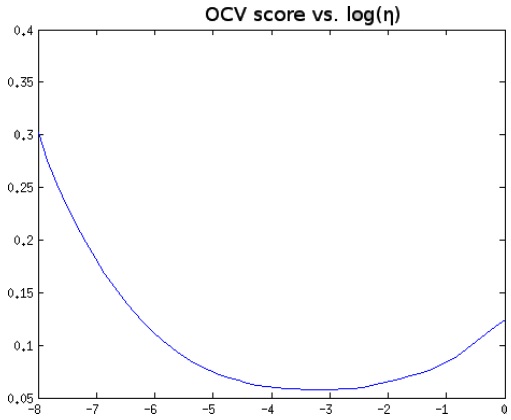


Figure 5.3: OCV plot vs $\log(\eta)$. The minimizer in this case is approximately $\eta = e^{-3}$.

5.2.6 Data Analysis

The quality of 3D reconstructions was gauged by calculating the contrast (C) and contrast-to-noise ratio (CNR) for pairs of regions in the TM phantom. In conventional 2D elastography, these statistics are calculated over carefully chosen regions of interest (ROI). For the 3D case, the same idea is extended here by choosing parallelepiped shaped ROIs, with lateral and elevation dimensions of 5mm each and 10mm axially. These statistics are presented in decibel units using the following formulas [144, 138]:

$$\mathrm{C} = 20\log_{10}\left(\frac{\mu_1}{\mu_2}\right)$$

and

$$\mathrm{CNR} = 20\log_{10}\left(\frac{2(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right).$$

where μ and σ respectively denote the mean and the standard deviation values of the SWVs calculated over each ROI and the subscripts indicate two distinct media. The mean and standard deviations over 5 independent datasets were obtained after converting to dB.

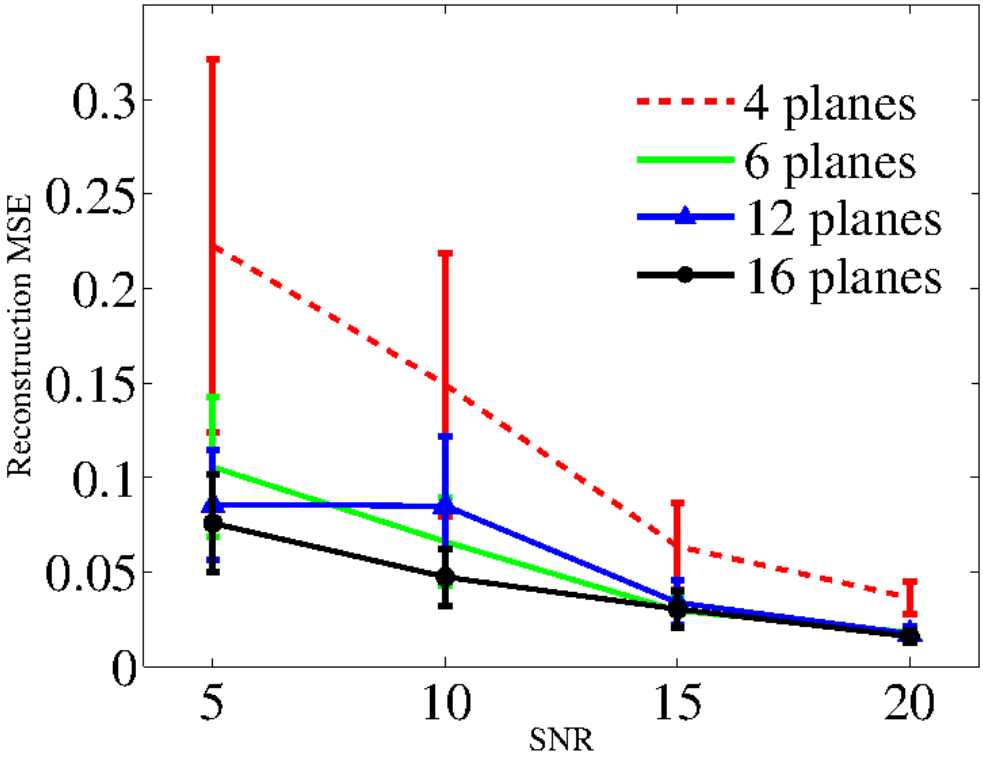


Figure 5.4: Mean squared reconstruction error (MSE) from simulated data at different noise levels and varying number of image planes. The reconstruction error is largest when only 4 image planes are used. No significant improvement is obtained by increasing the number of planes from 6 to 16, as seen from the overlapping errorbars around the MSE values.

Table 5.1: Shear wave velocity estimates SOUPR reconstruction

# imaging planes	background	partially ablated	inclusion
4	0.76 ± 0.07	1.01 ± 0.02	1.33 ± 0.10
6	0.75 ± 0.04	1.02 ± 0.02	1.34 ± 0.14
12	0.76 ± 0.06	0.99 ± 0.03	1.34 ± 0.10
16	0.76 ± 0.07	0.98 ± 0.03	1.33 ± 0.10
SSI	0.9 ± 0.07	1.1 ± 0.05	1.2 ± 0.03

Values of SWV (in m/s) for the three regions of interest are shown. The number of imaging planes used for 3D reconstruction is varied from 4 to 16. The mean and standard deviations were calculated over 3D parallelepiped shaped ROIs. For comparison, measurements from the commercial Supersonic Imagine (SSI) ShearWave^{TM} Elastography imaging mode are also shown. Note that the SSI measurements were obtained using conventional 2D ROIs from only a single imaging plane with the partially ablated region in view (cf. Fig. 5.9).

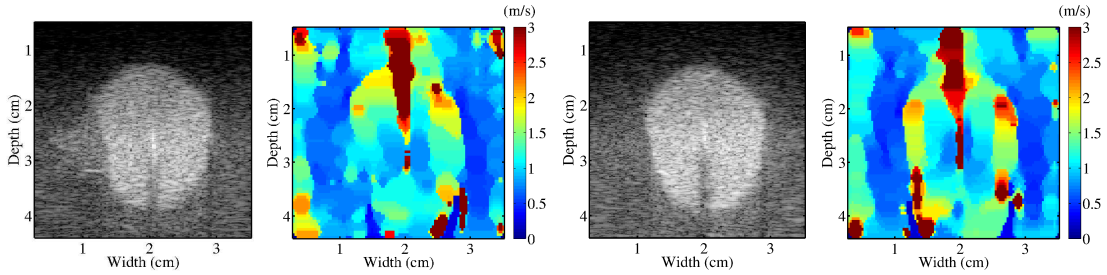
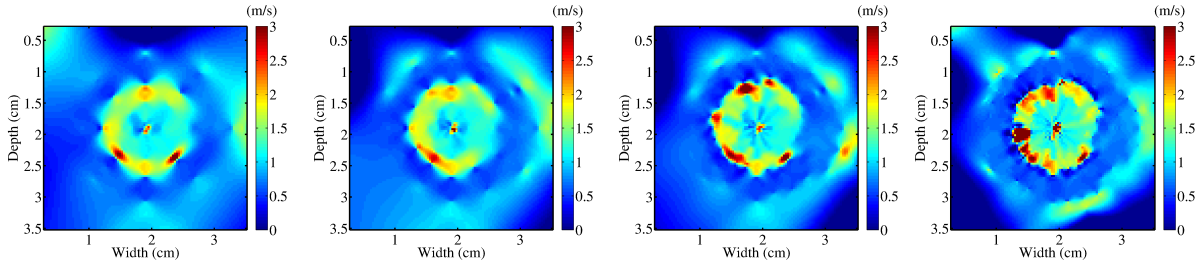
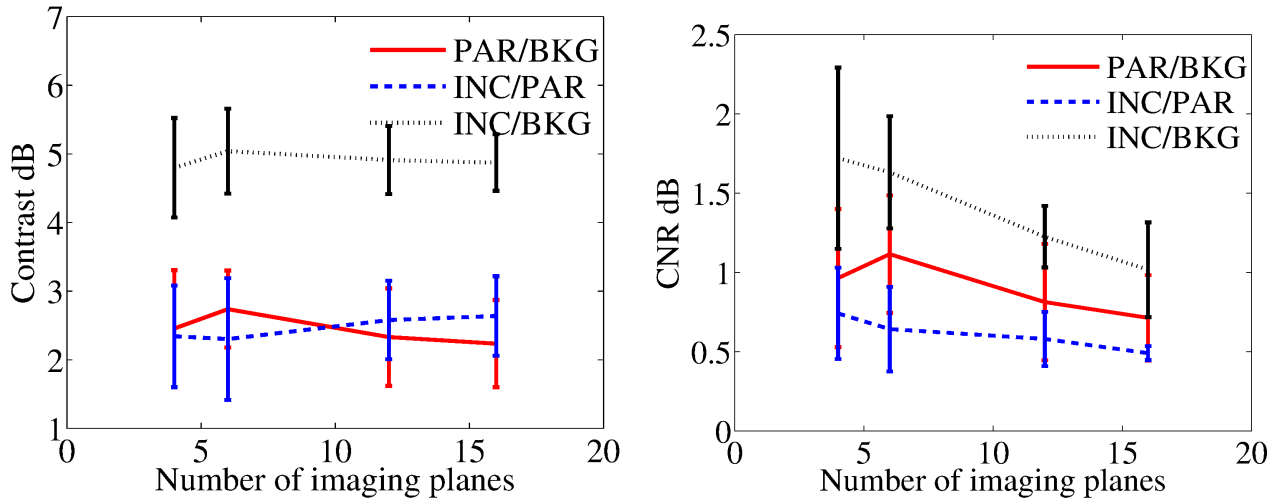


Figure 5.5: Ultrasound B-mode image and respective shear wave velocity (SWV) images are shown. Note the clear visualization of the partially ablated region seen on B-mode in (a) and in the shear wave image obtained using the piecewise linear fitting algorithm in (b). B-mode (c) and SWV reconstruction (d) for an image plane not including the partially ablated region are also shown for comparison. Backscattered intensity was varied in the phantom to visualize normal, ablated and partially ablated areas in the B-mode reference images.



(a) 4 imaging planes (b) 6 imaging planes (c) 12 imaging planes (d) 16 imaging planes

Figure 5.6: Example C-plane reconstructions of shear wave velocity at an axial depth of 2.8cm with different number of imaging planes. Observe that as more planes of data are introduced, the variability and finer detail in the boundary increases, and the "wheel spoke" artifact becomes less pronounced.



(a) Contrast (b) CNR

Figure 5.7: Plots of (a) Contrast and (b) CNR as a function of the number of imaging planes. Contrast and CNR plots are for three different pairs of regions. (BKG=background, PAR=partially ablated, INC=inclusion)

5.3 Results

5.3.1 Simulations

Results of Monte Carlo simulations are shown in Fig. 5.4. Note that the reconstruction error decreases with increasing SNR, irrespective of the number of image planes used.

In all cases the reconstruction error was largest when 4 image planes are used. On average, the 16 plane case gives the best reconstruction accuracy, but the error bars overlap considerably when the number of planes is 6, 12 or 16. This suggests that for the ideal case of a perfectly radially symmetric inclusion, increasing the number of image planes beyond 6 does not lead to any significant improvement in reconstruction quality. A more realistic scenario involves inclusions that are not perfectly ellipsoidal in shape. This is presented using TM phantom data where the phantom has an irregularly shaped region on one side.

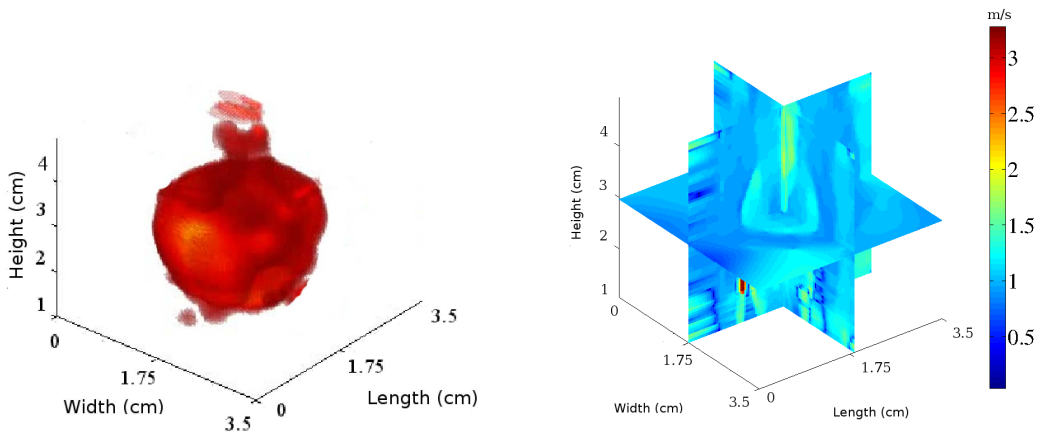
5.3.2 Phantom Data

The reconstructed SWV maps for two of the imaging planes over which RF ultrasound data loops were acquired are shown in Fig. 5.5. SWV reconstructions were performed by applying piecewise linear fitting to the raw TTP data as illustrated in Fig. 4.1. The first image plane passes through the partially ablated region which can be visualized in both the B-mode and the SWV reconstruction while the second plane is aligned such that the partially ablated region is not seen. Note that these regions of varying stiffness are not as easily visible in real tissue using a B-mode scan. But they can be easily seen here because the TM phantom material was designed to have different acoustic echogenicities for different regions. C-plane slices using different numbers of angular planes between 4 and 16 are shown in Fig. 5.6. The partially ablated region can be seen on the right side of the inclusion in all the four C-planes.

The estimated SWVs shown in Table 5.1 do not change significantly as the number of planes in the sheaf was changed. This is because the structure of the phantom is quite symmetrical, so on average the measured SWVs do not vary when new planes of data are added. The regularized function estimation framework described here is

also related to classical statistical learning theory: the choice of η is related to the so called bias-variance tradeoff [148, Ch. 2, Sec. 2.2.2]. A small value of η will produce an undulating fit, i.e., an estimator with low bias but high variance. On the other hand, larger value of η will reduce the variance and increase the bias. The use of Akaike information criterion in the piecewise linear fitting algorithm and regularization in the optimization problem are both ways to trade off some variance at the cost of introducing a bias in the estimated function. As a result of the smoothing, the SWV estimates shown in Table 5.1 are lower than the true values for the phantom calculated using the known material stiffness (Young's modulus). But they agree quite well with the SWV estimates obtained using SSI shown in the last row of the table. Moreover, delineation of ablated and partially ablated regions, which is of paramount importance, can be easily seen in the C-plane images.

There are several tradeoffs associated with increasing the number of imaging planes for 3D reconstruction. Additional planes would require significantly more data acquisition and processing to obtain SWV reconstructions. A quantitative evaluation of improvements obtained by increasing the number of imaging planes is therefore necessary to determine a reasonable number of planes needed for fast imaging. Image quality metrics calculated over 5 independent datasets are shown in Fig. 5.7. The largest contrast of about 5 dB in Fig. 5.7(a) is obtained when comparing the soft background with the hard inclusion, also imaged with 6 imaging planes. A highest CNR of 1.5 dB is obtained using 6 imaging planes in the sheaf as illustrated in Fig. 5.7(b). As one would expect, the highest contrast values were seen between the inclusion-background regions. In general, there is a decreasing trend in all image quality statistics as the number of planes was increased.



(a) 3D render (b) slice view

Figure 5.8: (a) 3D reconstructed volume of the inclusion using 6 imaging planes obtained by thresholding. (b) Three slice view of the 3D volume with SWV shown in m/s. (The irregularly shaped region is not visible in the planes shown here.)

The full 3D reconstruction using 6 imaging planes is shown in Fig. 5.8. This is created by stacking together the individual C-planes. This 3D render should be interpreted with caution because only an apparent estimate of the shear wave velocities is obtained. It should not be misconstrued as a solution of a full 3D shear wave propagation inverse problem. Fig. 5.8(a) is obtained by thresholding and is for visualization only. Numerical information is shown in the three-slice view in Fig. 5.8(b).

5.4 Discussion

This chapter presented a novel technique for acquiring, processing and displaying a 3D rendering of the SWV distributions for ultrasound shear wave elastography using a stack of C-plane reconstructions. The C-plane visualizations of the ablation slices can also be utilized to clearly delineate ablation boundaries. The sheaf of ultrasound

planes reconstruction (SOUPR) algorithm is presented and validated using TM phantom experiments for EVE.

A piecewise linear fit was used to accentuate change points with the goal of improving boundary delineation in the reconstructed SWV maps on individual scan planes. Other noise filtering methods such as least squares linear or polynomial fitting [65], Kalman filtering [149] or and other Bayesian model based filtering methods (Sections 4.4 and 4.5) can also be used. The idea of Tikhonov regularization for smooth function approximation was previously applied to displacement estimation by Rivaz *et al.* [150]. In contrast, SOUPR uses regularized optimization to reconstruct smooth shear wave velocity maps from undersampled grids. It is worth noting that the second order finite differencing regularization term used in this work can be replaced with first or higher order derivatives or any other types of penalty functions that promote smoothness. The key requirement is the ability to express the penalty as a linear operation so that the formulation in (5.1) can still be used.

The contrast stays almost constant as the number of planes in the sheaf was increased. The CNR quality metric shows a decreasing trend as the number of planes is increased. This may be a side effect of the smooth C-plane fitting algorithm. The fitting algorithm approximates a smooth surface in regions where there are no data points. With fewer number of planes there are larger regions with no data points, resulting in a visually smoother fit which corresponds to better CNR in the image. On the other hand, when the number of image planes is increased, the voids are much smaller. Therefore the first term in the optimization problem (5.1) exerts greater influence forcing the surface fit to appear more undulating. To enable fairer comparison between reconstructions performed with different number of image planes, these quality metrics may have to be modified by accounting for these effects.

A visible side effect of the sheaf imaging strategy is the "wheel spoke" artifact seen in Fig. 5.6, especially when using fewer planes in the sheaf. These appear as radial streaks emanating from the center of the C-plane and coinciding with the radial locations of the image planes over which ultrasound echo data was acquired. The smoothing parameter can be manually tuned to get rid of such artifacts—a larger value of η will smooth out such variations. In a commercial system, there may be regulatory hurdles to allowing the user to control the value of η, but it may be possible to include an additional preset akin to smoothing and edge enhancement which is available on almost all ultrasound scanners.

In a real world ablation monitoring situation, it may be more challenging to align the needle along the axial beam direction of the transducer. In situations where the needle is at an angle to the axial ultrasound beam direction, 2D displacement tracking algorithm can be used. The component of these displacements parallel to the needle can be extracted and TTP values along lines perpendicular to the needle can be calculated in steps 2 and 3 of Section 5.2, respectively. In case of physical constraints on placing the transducer around the needle, it may be impossible to obtain the complete sheaf, resulting in voids in the dataset. But a partial volume reconstruction can still be obtained with a SOUPR-like algorithm, using only that part of the fine grid where data points are present. The four step decomposition offers good flexibility for tuning the details of each signal processing block for specific applications. For example, some commercial ultrasound scanners already offer imaging modes to map shear wave velocities (either as point estimates or over large regions) irrespective of the alignment of the image plane with respect to the needle. In such systems, it will be convenient to incorporate one additional block for step 4 at the end of the existing signal processing chain to generate 3D volume maps.

It is also useful to compare the SWV image plane reconstruction with a "bronze standard" image acquired from a commercial Supersonic Imagine ultrasound scanner with shear wave imaging capability. Fig. 5.9 shows one image plane of the phantom with a SWV overlay on B-mode. The image plane is identical to the one shown in Figs. 5.5(a) and (b). Note that high velocity artifacts can be seen at depths greater than 2.5cm.

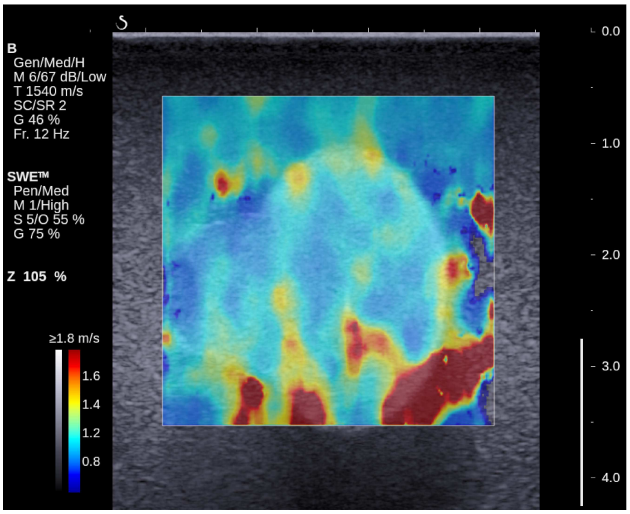


Figure 5.9: A "bronze standard" SWV image of the phantom inclusion corresponding to the same imaging plane of Fig. 5.5(a). This image was acquired using a Supersonic Imagine ultrasound scanner with the shear wave imaging mode. (Original image has been cropped and rearranged to fit.)

The results presented here indicate that 6 planes in a sheaf are sufficient for fast reconstruction in the particular TM phantom used, both qualitatively and quantitatively. This is because the shape of the inclusion was mostly symmetric about the needle, (except for a small irregularly shaped partially ablated region). This was also confirmed by the simulation experiment that was specifically used to mimic the situation in the phantom experiment. Greater number of image planes may be needed if the inclusion is more irregularly shaped. More complex geometries can be simulated ahead of time

using the method described in Section 5.2.5 to decide on a suitable number of image planes in the sheaf as part of pre-ablation clinical planning.

The location and number of angular planes can also be optimized with the aim of minimizing the number of planes required for effective 3D reconstruction while maximizing the processing speed. This will also allow implementation of 3D reconstruction algorithms on low end commercial ultrasound systems which may not be equipped with sufficient processing power or modern graphical processing units (GPUs). Minimization of the number of planes will enable faster data acquisition and processing for eventual real-time implementation of 3D SWV volume reconstructions. Adaptive or non-uniform sampling of complex regions, for example those including the simulated, partially ablated regions as shown in Figs. 5.5–5.6 and regions near large vessels during *in vivo* implementations, will have to be evaluated to determine if this can further improve delineation. Data can be acquired in multiple passes where the interpolated visualization from an earlier pass provides feedback for sampling critical locations in the volume, enabling an adaptive sampling approach for improved delineation.

Although the sheaf pattern of acquisition is naturally suited to the geometry of the EVE setup, an inherent limitation of this method is that every image plane must contain the needle to enable time of arrival estimation. In a sheaf, data samples progressively get sparser away from the needle. This shortcoming can be addressed by sampling extra planes in the sheaf so that a certain fine sample spacing is achieved at a predetermined distance away from the needle. Alternatively, if a specific ROI has already been located, imaging planes that are not uniformly spaced in angle can be employed; finer spacing in specific regions may provide better reconstruction quality at those locations. Knowledge from earlier iterations of the reconstruction algorithm may also be incorporated in the interpolation procedure, by reconstructing only parts of the full volume that are known

to contain any interesting features, either automatically or with the intervention of a clinician.

6

Kernel Based Methods for 3D Reconstruction

> Essentially, all models are wrong, but some are useful.
>
> G.E.P. Box, N.R. Draper, *Empirical Model Building and Response Surfaces*

6.1 Introduction

Real time three dimensional visualization of percutaneous needle ablation procedures has considerable clinical value. Traditionally, 1D linear or phased array transducers have been used to reconstruct strain and/or shear stiffness to delineate the ablation boundary in 2D imaging planes. However, ablation is fundamentally a 3D process and hence a 3D imaging modality is desirable. Although matrix transducers can be used to obtain full volume acquisitions, they provide lower frame rates and are more expensive than 1D arrays.

A prototypical problem in data visualization is that of scattered data interpolation. In shear wave elastography, stiffness measurements may be available only at a small subset of points in the imaged volume. For instance, when using the sheaf of ultrasound planes based reconstruction (SOUPR), data is only available on a few image planes in a

spoke-wheel pattern. In case of 3D volumes acquired with a matrix array, data may be only sparsely sampled in the entire volume in the interest of faster frame rates. In both situations, it is necessary to reconstruct the complete shear wave velocity map over a fine grid of points in a complete volume. This is a challenging problem because the problem will typically be ill-posed due to the availability of few data points compared to the size of the grid, i.e., more unknowns than known data points.

This chapter introduces a Matérn kernel based reconstruction method which provides an objective way of imposing "prior structure" on the final reconstruction. An oversimplified model would assume that there is an abrupt variation in stiffness going from the ablated region to the surrounding unablated tissue. However, in reality, previous work has shown that there is a transition region between the stiff and softer background. Moreover, this transition may be very difficult to describe mathematically because the process of ablation involves a complex interplay between a model for transfer of heat into tissue and a model of cell necrosis as a function of temperature and time. The Matérn model provides an adaptive method for learning the degree of smoothness of the final reconstruction from underlying data. The use of Matérn kernel with carefully selected parameters allows fast reconstruction from undersampled noisy data volumes while maintaining stiffness boundary details.

This chapter also analyzes some theoretical properties of the proposed algorithm in the context of reproducing kernel Hilbert spaces. It can be shown that the smoothing algorithm produces optimal reconstructions in a predictive mean squared error sense. This is corroborated using simulation and tissue-mimicking (TM) phantom-based experimental results which show that this method in fact reduces spoke wheel artifacts seen in SOUPR.

6.2 Reproducing Kernel Hilbert Space (RKHS) Formalism

Restricting attention to individual transverse planes, the SWV field can be represented by a function $f : \mathbb{R}^2 \rightarrow [0, \infty)$. With a slight abuse of notation, t is used in this chapter for locations of SWV measurements in the imaged volume. Noisy point evaluations of this function are acquired at different locations $\{t_i\}_{i=1}^n$.

$$u_i = f(t_i) + \epsilon_i$$

where each $t_i \in \mathbb{R}^2$ is situated along radial lines and ϵ_i is assumed to be i.i.d. Gaussian distributed noise with unknown variance. A reconstruction of f (via interpolation or approximation) on a fine grid of M points is desired. As it stands, the problem is ill-posed because the number of data points n is usually much smaller than the size of the reconstruction grid ($n \ll M$). Typically, the number of data points obtained on each plane is on the order of 10^3, whereas the grid may consist of $10^4 - 10^5$ points. It is necessary to restrict the space of functions that the reconstruction algorithm operates on in order to guarantee good theoretical properties [151]:

Definition 6.1 (Reproducing Kernel Hilbert Space) A reproducing kernel Hilbert space $\mathcal{H}$ of functions f on $\mathbb{R}^2$ is defined as the Hilbert space of functions over which the point evaluation functional is a bounded linear functional, i.e., for every $t \in \mathbb{R}^2$ there exists $M_t < \infty$ such that $|f(t)| \leq M_t ||f||_{\mathcal{H}}$ for all f and $|| \cdot ||_{\mathcal{H}}$ denotes the RKHS norm.

In the most general setting, the function f is assumed to be a member of an RKHS $\mathcal{H}_R$ with the reproducing kernel R. So the optimization problem of interest is [152]:

$$\text{minimize} \sum_{i=1}^{n} (u_i - f(t_i))^2 + \lambda ||f||^2_{\mathcal{H}_R}. \qquad [6.1]$$

Here $f \in \mathcal{H}_R$ and $||\cdot||_{\mathcal{H}_R}$ denotes the RKHS norm. Another interpretation is to consider this as a regularized solution to an ill-posed inverse problem also called the Fredholm equation of the first kind, with the point evaluation function playing the role of the "blurring function" [153] in this case. The following theorem characterizes the existence of a solution to this optimization problem:

Theorem 6.2 (Representer Theorem) The solution of (6.1) can be expressed in the form:

$$f_\lambda(t) = \sum_{i=1}^{n} c_i R(t, t_i).$$

A short proof is presented in C.5. In fact, the values of c_i can be evaluated in closed form [154, 146]. This theorem has important practical ramifications because it reduces the problem of finding f in an infinite dimensional Hilbert space to a problem of finding coefficients c_i which is a finite dimensional problem. The smoothing parameter λ is chosen through generalized cross-validation (GCV) [146].

6.3 Matérn Model

The choice of kernel is paramount because it dictates the smoothness properties of the final fit. The Matérn radial basis function (RBF) can be used as a more general kernel function compared to the Gaussian RBFs. As it will be seen, the Matérn RBF reduces to a Gaussian RBF for certain choice of parameters. The Matérn RBF is defined as [155, ch. 2, sec. 2.7], [156]:

$$R(||t - t_i||) = \frac{\sqrt{\pi}}{2^{\nu-1}\Gamma(\nu + 1/2)r^{-2\nu}} \left(\frac{||t - t_i||}{r}\right)^{\nu} K_\nu \left(\frac{||t - t_i||}{r}\right) \qquad [6.2]$$

with $\nu, r > 0$ and K_ν is the modified Bessel function of the second kind. The unknown function can be approximated as:

$$\widehat{f}(t) = \sum_{i=1}^{n} c_i R(||t - t_i||).$$

Intuitively, this formulation suggests that each data point t_i imposes a "region of influence" that varies with its distance from an arbitrary point t. Note that the absolute locations of the points t and t_i are immaterial, only the distance $||t - t_i||$ plays a role in the evaluation of the kernel. The characteristics of this region of influence depend on the shape of R, which is controlled using two parameters ν and r. The smoothness (differentiability) is controlled using ν whereas the range of influence is controlled by r. It is customary to call ν the smoothness parameter whereas r is called the range (or scale) parameter of the Matérn covariance function. In particular, as $\nu \to \infty$ the Matérn Kernel converges to an infinitely differentiable Gaussian kernel function. The advantage of using the Matérn kernel is that it provides sharper roll-offs than the infinitely differentiable Gaussian function, and imparting the ability preserve boundary details. A common choice is to set $\nu = m + \frac{1}{2}$ for different non-negative integers m, resulting in a simple closed form expression for $R(\cdot)$ [155].

It can be shown that linear combinations of the Matérn kernel function produce functions whose frequency response decays like a reciprocal polynomial [157], where the rate of decay is related to the ν parameter. This is known as a Sobolev space formally defined as follows [158, Ch. 8]:

Definition 6.3 (Sobolev Space of functions on $\mathbb{R}^2$) The Sobolev space $\mathcal{W}^{\alpha}_{s,2}(\mathbb{R}^2)$ is defined as the space of square integrable functions f on $\mathbb{R}^2$ that satisfy

$$\int_{\mathbb{R}^2} (\alpha^2 + \omega_1^2 + \omega_2^2)^s |(\mathcal{F}f)(\omega_1, \omega_2)|^2 \, d\omega_1 \, d\omega_2 < \infty. \qquad [6.3]$$

Input: $\{t_i\}_{i=1}^n$: data point locations in $\mathbb{R}^2$
$\{u_i\}_{i=1}^n$: corresponding noisy data points
$\mathcal{S}_1$: finite set of smoothness values
$\mathcal{S}_2$: finite set of scale (range) values
Output: $f^*(\cdot)$: smooth function fit to the noisy data
R^*: Matérn kernel for the space that f^* belongs to
λ^*: smoothing parameter chosen using GCV

procedure MATÉRNSMOOTH($\{t_i\}_{i=1}^n, \{u_i\}_{i=1}^n, \mathcal{S}_1, \mathcal{S}_2$)
  $V^* \leftarrow \infty$
  for each $(\nu, r) \in \mathcal{S}_1 \times \mathcal{S}_2$ **do**
    $\lambda_{GCV} \leftarrow$ smoothing parameter chosen using GCV
    $V \leftarrow$ GCV score at λ_{GCV}
    $[\mathbf{\Sigma}]_{ij} \leftarrow R_{(\nu,r)}(||t_i - t_j||)$
    $\mathbf{c} \leftarrow (\mathbf{\Sigma} + \lambda_{GCV}\mathbf{I})^{-1}[u_1 u_2 \cdots u_n]^T$
    if $V \leq V^*$ **then**
      $V^* \leftarrow V$
      $f^*(\cdot) \leftarrow \sum_{i=1}^n c_i R_{(\nu,r)}(||\cdot - t_i||)$
      $R^* \leftarrow R_{(\nu,r)}$
      $\lambda^* \leftarrow \lambda_{GCV}$
    end if
  end for
  return (f, R^*, λ^*)
end procedure

Figure 6.1: Matérn kernel based reconstruction algorithm.

where

$$(\mathcal{F}f)(\omega_1, \omega_2) = \int_{\mathbb{R}^2} f(x_1, x_2) e^{-j(\omega_1 x_1 + \omega_2 x_2)}\, dx_1\, dx_2$$

denotes the 2D Fourier transform of f.

Furthermore, it can be shown that $\mathcal{W}_{s,2}^{\alpha}$ is a RKHS with the Matérn function as its reproducing kernel and hence the RKHS norm can be defined using the integral in (6.3).

Theorem 6.4 Let $R(\cdot)$ be as defined in (6.2). Then $\frac{\Gamma(\nu+\frac{1}{2})}{4\pi^{3/2}\Gamma(\nu+1)} R(\cdot)$ is the reproducing kernel for $\mathcal{W}_{(\nu+1),2}^{1/r}(\mathbb{R}^2)$.

The proof involves some algebra to calculate the inverse 2D Fourier transform of $(1/r^2 + \omega_1^2 + \omega_2^2)^{-\nu-1}$ as shown in C.6. Note that different authors set the leading normalization constants in different ways. For simulations and data analysis, the leading constant C was chosen in such a way that $C(d/r)^\nu K_\nu(d/r)$ is normalized to 1 when $d = 0$ for fixed (ν, r). This is made possible by the fact that $K_\nu(u)$ behaves like $2^{\nu-1}\Gamma(\nu)u^{-\nu}$ as $u \downarrow 0$. [159, Sec. 9.6.9, pp. 375].

The optimization problem (6.1) can now be solved over different spaces $\mathcal{H}_{R(\nu,r)}$ by choice of kernel parameters (ν, r). This is main idea behind the MatérnSmooth algorithm presented in Fig. 6.1. For a fixed pair of kernel parameters, choosing the smoothing parameter λ through GCV has the following desirable theoretical property:

Theorem 6.5 (GCV Theorem) Let f_λ be the solution of the optimization problem in (6.1) and define the predictive mean squared error as $\mathsf{PMSE}_n(\lambda) = \frac{1}{n}\sum_{i=1}^n (f_\lambda(t_i) - f(t_i))^2$. Let λ_{opt} be the optimal value of the smoothing parameter (that depends on the unknown function f) that minimizes the PMSE and let λ_{GCV} be the value chosen using generalized crossvalidation. Also define the inefficiency function $I(n) = \frac{\mathsf{PMSE}_n(\lambda_{GCV})}{\mathsf{PMSE}_n(\lambda_{opt})}$. Then as the number of data points $n \uparrow \infty$, $I(n) \downarrow 1$.

A proof of this theorem can be found in the paper by Utreras [160] and will not be reproduced here.

Corollary 6.6 Let the hypotheses of Theorem 6.5 hold. Then the algorithm in Fig. 6.1 has the property that as $n \uparrow \infty$, $I(n) \downarrow 1$ on the union of Sobolev spaces $\bigcup_{(\nu,r)\in\mathcal{S}_1\times\mathcal{S}_2} \mathcal{W}^r_{\nu,2}$.

The corollary is proved in C.7.

6.4 Results

6.4.1 Simulations

The following function definition was used for generating simulated 3D data:

$$g(x,y,z) = \begin{cases} 4 & \text{if } x^2 + y^2 + \frac{(z-2.25)^2}{1.5^2} \leq 1 \\ 1 & \text{otherwise.} \end{cases}$$

This mimics a stiff inclusion in a soft background with a stiffness ratio of 4, with the center of the inclusion at a depth of 2.25 cm below the surface of the phantom as shown in Fig. 6.2. Data was generated at different depths z and the reconstruction algorithm was applied over each plane at constant depth.

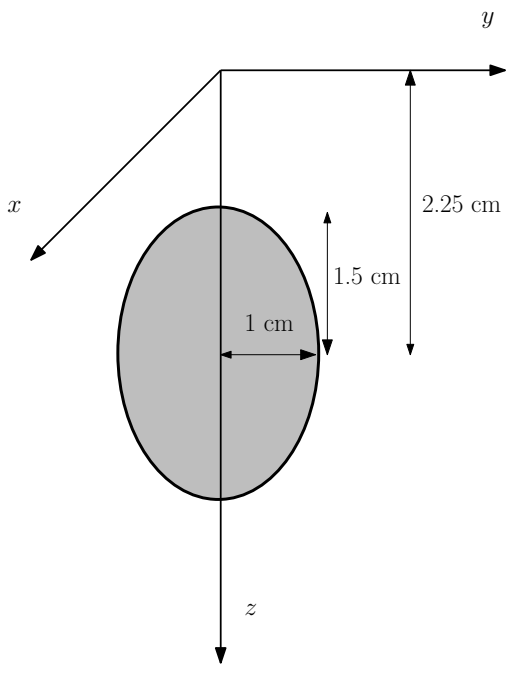


Figure 6.2: Function model used for generating simulated data. The stiffer region is shaded gray and is surrounded by softer region. The dimensions match the design of an actual phantom that is used later for experimental evaluation of the algorithm.

To guide the selection of kernel parameters, this data generator was used to simulate noisy data with a radial acquisition pattern. Three different cases for $P = 6, 12, 16$ using angular increments of π/P, one hundred radial samples over 4 cm and 100 samples over a depth of 4.5 cm in the z direction were used. The reconstruction was performed in the bounded box $[-2, 2] \times [-2, 2] \times [0, 4.5]$. Zero mean additive Gaussian noise with a standard deviation of $\sigma = 0.5$ was used, and the simulation was run 10 times for each pair of (ν, r) values, with $\nu \in \{0.2, 0.9, 1.5, 2.5\}$ and $r \in \{0.04, 0.1, 0.2, 0.4, 0.6\}$. Note that the standard deviation of 0.5 when detecting a jump of 3 units corresponds to a signal to noise ratio of approximately 15 dB which resembles the kind of noise seen in ultrasound based SWV measurements. The numerical values used in this simulation were chosen to match the ground truth that one theoretically expects to observe on a phantom with an ellipsoidal inclusion.

The SWV is reconstructed on a fine grid of $M = 10^6$ points using the Matérn radial basis function approximation and the predictive mean squared error is estimated using:

$$\mathsf{PMSE}(\nu, r) \quad = \quad \frac{1}{M} \sum_{l,m,n} \left[\widehat{g} \left(\frac{4(l-50)}{100}, \frac{4(m-50)}{100}, \frac{4.5n}{100} \right) - g \left(\frac{4(l-50)}{100}, \frac{4(m-50)}{100}, \frac{4.5n}{100} \right) \right]^2 .$$

where $\hat{g}$ is the reconstructed function, g is the true function and the sum is calculated over a discrete grid of points $0 \leq l, m, n \leq 99$ and $(l, m, n) \in \mathbb{Z}^3$. Mean and standard deviation PMSE from these Monte Carlo simulations are shown in Tables 6.1, 6.2 and 6.3.

It is clear from these tables that the PMSE improves as the number of radial lines P increases. The pair $(\nu, r) = (0.9, 0.4)$ gives the lowest reconstruction error in all three cases. The next best choice of parameters, if first order mean squared differentiability is desired, are $(\nu, r) = (1.5, 0.4)$. This choice of kernel parameters is used later for reconstructions from experimental data acquired on a TM phantom. It is worth noting that

Table 6.1: Mean and standard deviation of PMSE for $P = 6$ radial lines

r \ ν	0.2	0.9	1.5	2.5
0.04	0.920 (0.019)	0.650 (0.015)	0.551 (0.013)	0.449 (0.011)
0.1	0.546 (0.020)	0.276 (0.011)	0.197 (0.009)	0.136 (0.007)
0.2	0.284 (0.017)	0.117 (0.007)	0.092 (0.006)	0.089 (0.008)
0.4	0.154 (0.011)	0.085 (0.006)	0.088 (0.010)	0.149 (0.041)
0.6	0.126 (0.009)	0.085 (0.007)	0.099 (0.016)	0.264 (0.098)

Table 6.2: Mean and standard deviation of PMSE for $P = 12$ radial lines

r \ ν	0.2	0.9	1.5	2.5
0.04	0.778 (0.015)	0.420 (0.011)	0.313 (0.009)	0.221 (0.007)
0.1	0.341 (0.013)	0.134 (0.006)	0.102 (0.005)	0.087 (0.004)
0.2	0.167 (0.008)	0.084 (0.004)	0.078 (0.004)	0.080 (0.003)
0.4	0.108 (0.005)	0.074 (0.003)	0.078 (0.003)	0.114 (0.011)
0.6	0.096 (0.005)	0.074 (0.003)	0.083 (0.004)	0.167 (0.027)

Table 6.3: Mean and standard deviation of PMSE for $P = 16$ radial lines

ν \ r	0.2	0.9	1.5	2.5
0.04	0.688 (0.022)	0.317 (0.014)	0.223 (0.011)	0.153 (0.008)
0.1	0.265 (0.015)	0.107 (0.006)	0.087 (0.005)	0.079 (0.004)
0.2	0.135 (0.009)	0.075 (0.004)	0.072 (0.003)	0.077 (0.003)
0.4	0.093 (0.006)	0.068 (0.003)	0.074 (0.003)	0.128 (0.021)
0.6	0.084 (0.005)	0.069 (0.003)	0.082 (0.004)	0.206 (0.055)

the reconstruction quality is more sensitive to the choice of range r than the smoothness ν. A surface plot of the PMSE over a wider range of ν and r values is shown in Fig. 6.3. By controlling ν and r it is possible to reduce the spoke-wheel artifacts that appear due to the radial sampling strategy (Fig. 6.4).

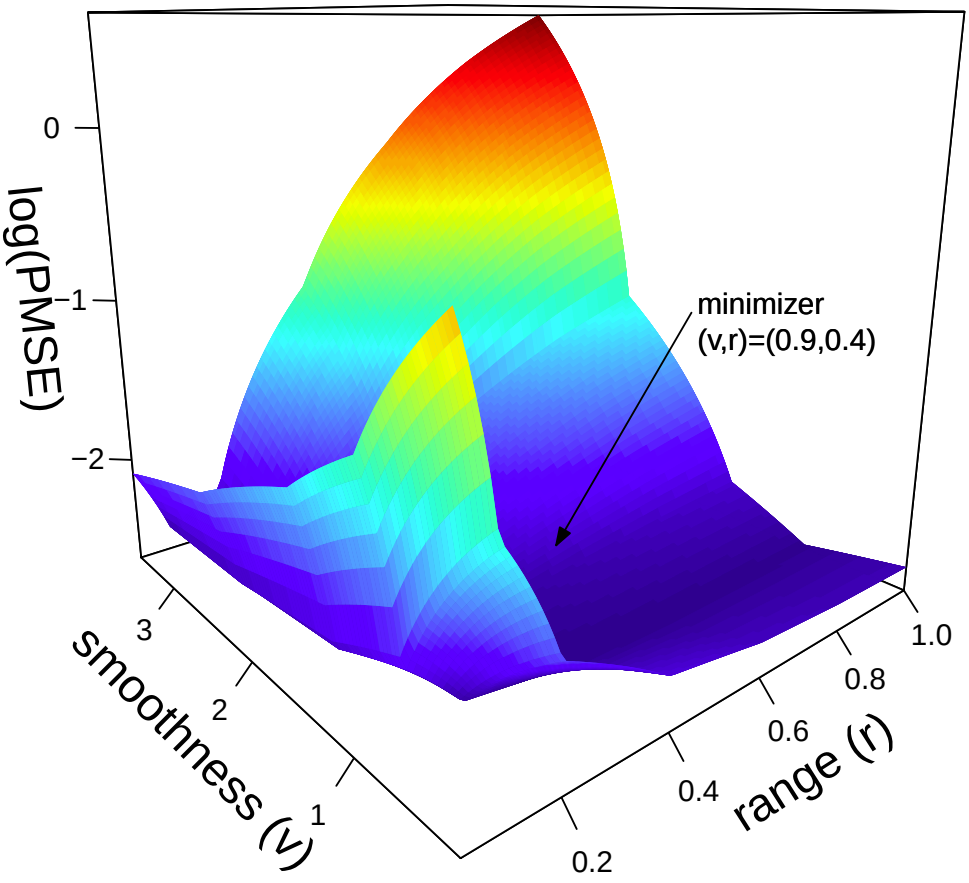
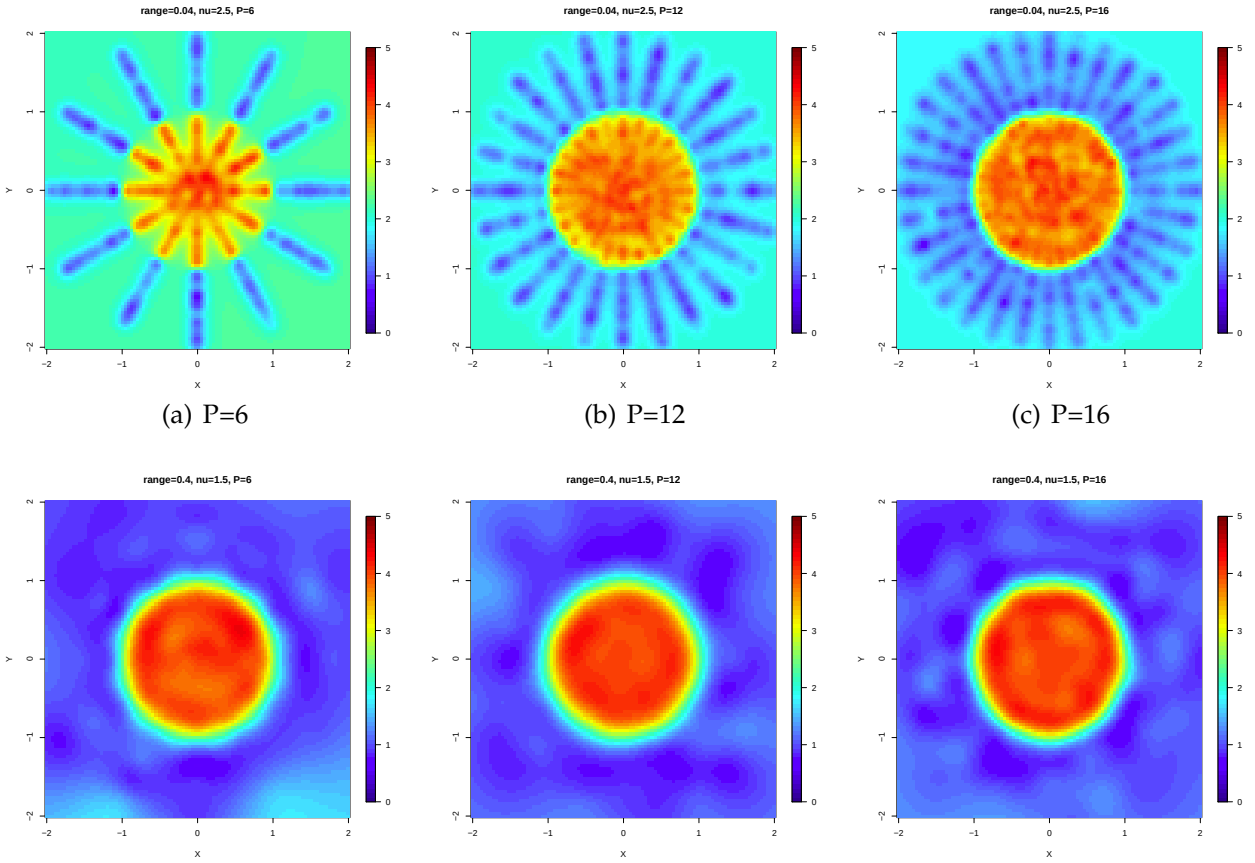


Figure 6.3: PMSE values for different combinations of (ν, r). The local minimizer is also shown. For every pair of (ν, r) values, the tuning parameter λ was chosen through GCV.



(a) P=6 (b) P=12 (c) P=16

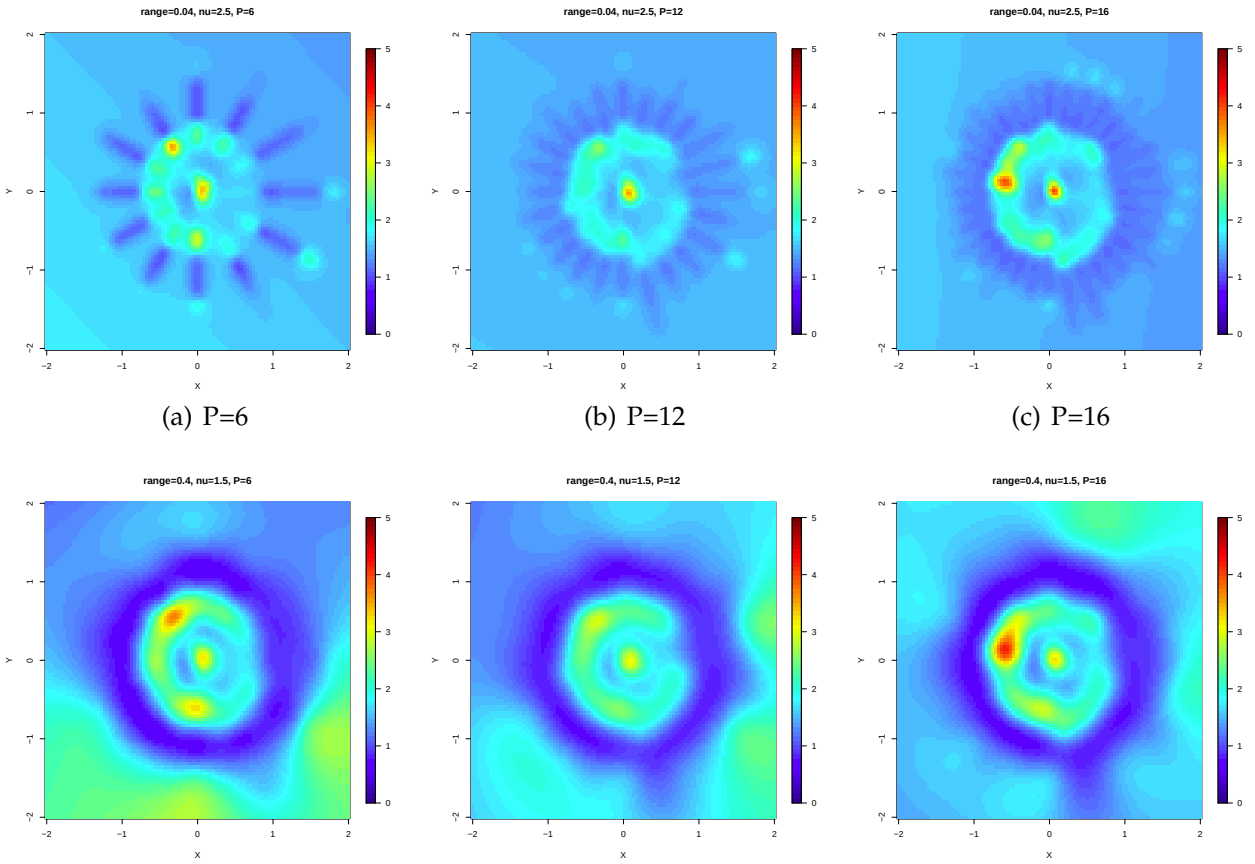
(d) P=6 (e) P=12 (f) P=16

Figure 6.4: Reconstructions from simulated data using different Matérn kernel parameters for different numbers of radial lines of data. Top row shows the "spokewheel" image artifact which can be avoided by proper choice of (ν, r) as seen in the bottom row. Top row $(\nu, r) = (2.5, 0.04)$ and bottom row $(\nu, r) = (1.5, 0.4)$.

6.4.2 Experimental Results

Experimental data was acquired on a TM phantom that consisted of an ellipsoidal inclusion in a soft background with known stiffness. The approximate mathematical description of the size and location of the inclusion matches the model used for generating simulated data in Section 6.4.1. Three different data sets are obtained experimentally by changing the number of radial sampling lines to $P = 6, 12$ and 16, respectively.

Fig. 6.5 shows the reconstructed SWV images over one particular image plane passing through the center of the inclusion. The top row of images show reconstructions using incorrectly chosen kernel parameters which result in bright streaks along the radial sampling lines. But when the "optimal" (ν, r) pair is used, this spokewheel artifact is greatly reduced as seen in the second row of images.



(a) P=6 (b) P=12 (c) P=16

(d) P=6 (e) P=12 (f) P=16

Figure 6.5: Reconstructions of experimental data using different Matérn kernel parameters for different numbers of radial lines of data. Top row shows the "spokewheel" image artifact is ameliorated by using $(\nu, r) = (1.5, 0.4)$ as seen in the bottom row. Top row $(\nu, r) = (2.5, 0.04)$ and bottom row $(\nu, r) = (1.5, 0.4)$.

Inclusion volume is an important quantity that helps clinicians gauge the extent of the ablated region and determine if additional treatment would be necessary. A clinician usually measures the principal axes of the inclusion from the reconstructed images displayed on a graphical user interface. Volume estimates for the ellipsoid are obtained from experimental reconstructions by measuring the three semi-principal axes (a, b, c) and using the ellipsoid volume formula $V = \frac{4}{3}\pi abc$. Results for different P along with the ground truth volume are shown in the box plot in Fig. 6.6.

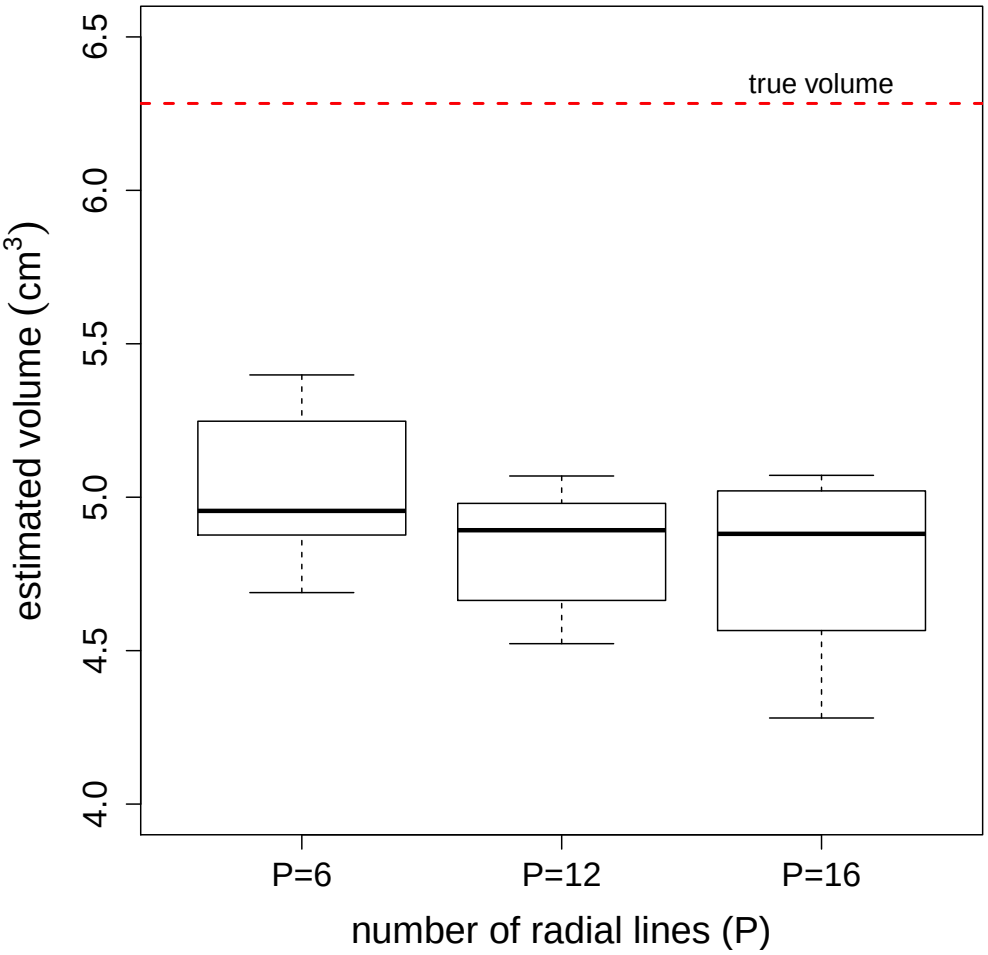
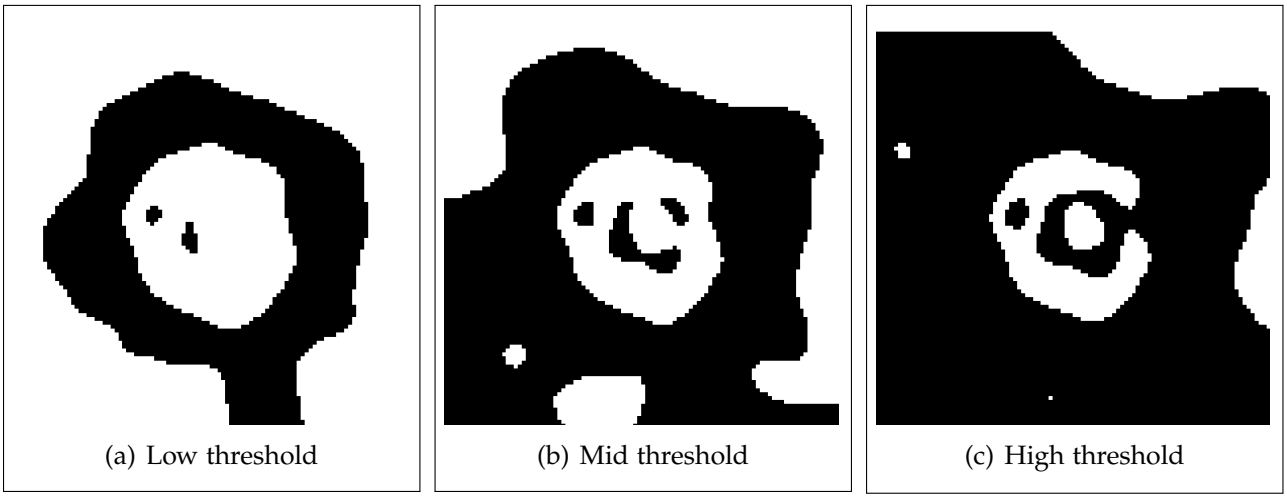


Figure 6.6: Box plot showing the volume estimates when the ellipsoidal inclusion is reconstructed using $P = 6, 12, 16$ radial acquisition lines. The true volume is shown using the dashed horizontal line.

The estimated volume is biased and underestimates the true volume in all three cases by about 20%. This is expected because there are multiple smoothing and noise filtering steps in the signal processing leading up to the final images shown in Fig. 6.5, which erodes the boundary detail. The bias in measuring the principal axes of the ellipsoid is only about 7% which gets magnified by a factor of 3 when calculating the volume. It is possible to create a thresholding procedure that converts the reconstructed images into binary images that are white inside the inclusion and black outside. The clinician can control a slider to smoothly transition between different levels of thresholding and obtain images like Fig. 6.7.



(a) Low threshold (b) Mid threshold (c) High threshold

Figure 6.7: Thresholded images created from the $P = 16$ image from Fig. 6.5(f). As the threshold is gradually increased, a smaller region of the image appears as an inclusion.

6.5 Discussion

The Matérn kernel reconstruction algorithm developed in this chapter provides an objective method for choosing the degree of smoothness for 3D reconstructions. This

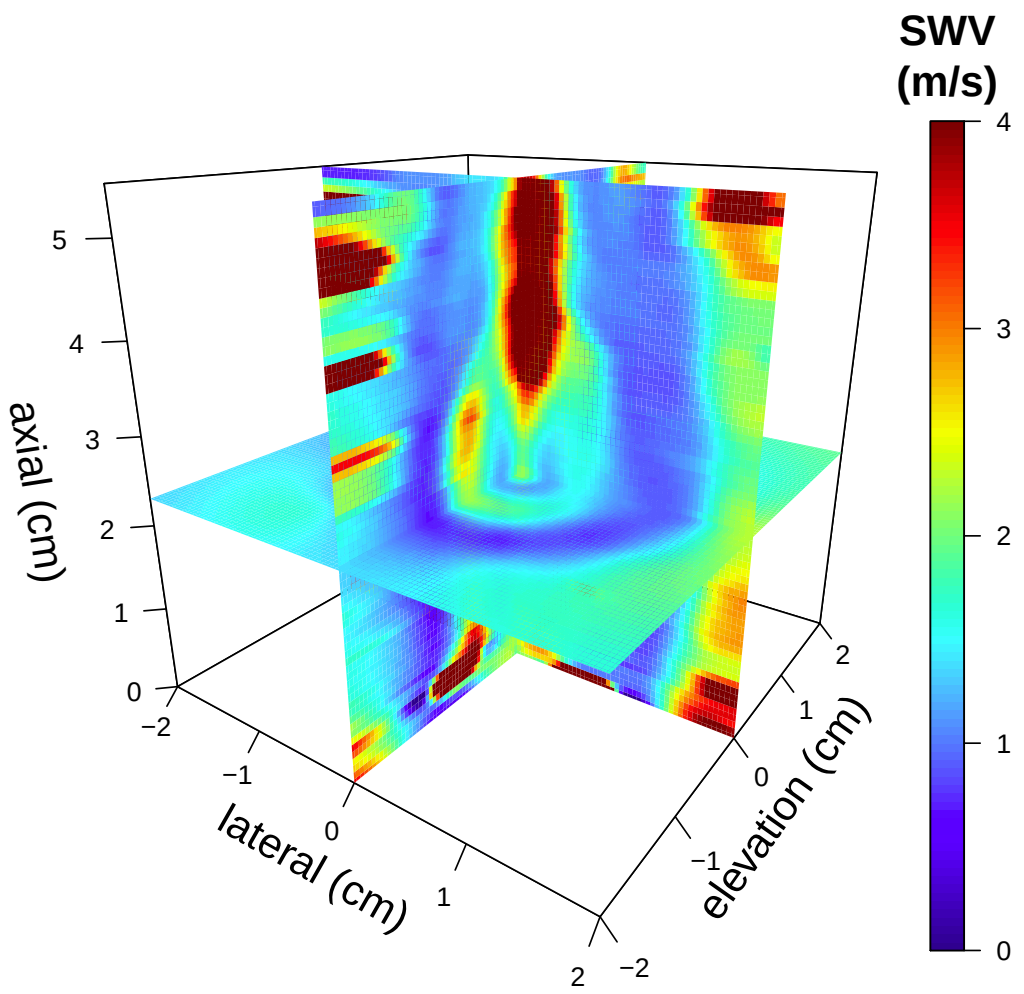


Figure 6.8: A 3D render of the stiff inclusion phantom produced from a stack of individual 2D planes at different depths.

has practical significance because the process of ablation is governed by a complex coupled system of heat transfer and tissue necrosis, that may be difficult to model accurately [161]. The algorithm presented in this paper has good theoretical properties in that it applies an "optimum" amount of smoothing by careful choice of the smoothing parameter and the kernel function. The final output will be sensitive to the choice of the sets

$\mathcal{S}_1$ and $\mathcal{S}_2$. Although it is difficult to develop a general theory on choosing these sets, it is possible to follow some heuristic guidelines. By inspecting the B-mode image of the ablated area, the clinician can get an idea of the sharpness of the transition from ablated region to healthy tissue. Their subjective judgment can be used to guess some reasonable range of integer values for degree of smoothness for the transition and choosing $\nu = m + 1/2$ for small integer values of m. Additionally, if the B-mode image suggests any discontinuities, values of $\nu < 0.5$ that will promote even sharper, non-differentiable reconstructions may be used. Unlike ν which has an intuitive relationship to the smoothness of the final fit, the choice of scale parameter r is more subtle. As a rule of thumb, r should be large enough so that more than one neighboring data point is captured by a kernel centered around any data point, and should not be so large that far away points are unnecessarily weighted heavily. Different values of r corresponding to the k^{th} smallest distance between any two data points can be tried for, say, $k = 1, 2, 3, \ldots$, and using the simulation method discussed previously, the (ν, r) values that provide a local minimum value of PMSE should be used.

7

Full Volume Reconstruction using a Markov Random Field Model

> "I like to point out that the model really should be called Lenz-Ising model. My supervisor, Dr. Wilhelm Lenz, had the idea and proposed that I make a mathematical workout as my dissertation, . . . "
>
> Ernst Ising in a letter to S. Kobe, *Braz. J. Phys. Dec. 2000*

7.1 Introduction

This chapter presents an efficient 3D reconstruction algorithm for ultrasound elastography data using a sparse iteration based on a Markov random field model. This method extends the previously reported sheaf reconstruction method to generate complete three dimensional visualizations of shear wave velocities without resorting to an approximate intermediate step of reconstructing transverse C-planes. The noisy data is modeled using a Markov random field, and a computationally tractable reconstruction

algorithm that can handle grids with millions of points is developed. Results from simulated ellipsoidal inclusion data show that this algorithm outperforms standard nearest neighbor interpolation by an order of magnitude in mean squared reconstruction error. Results from phantom experiments show that it also provides a higher contrast-to-noise ratio by almost 2 dB and better SNR in the stiff inclusion by over 2 dB.

The sheaf of ultrasound planes reconstruction (SOUPR) method was presented in [162] for obtaining C-plane (i.e., transverse plane) reconstructions of SWVs. Instead of using an approximate approach of reconstructing individual transverse planes, this chapter presents a more general method for reconstruction on a full 3D grid while also handling ill-posed situations where the number of grid points may exceed the number of data points. A computationally tractable algorithm that can handle large grid sizes with millions of grid points in 3D is also presented. A standard method for dealing with fast interpolation of scattered data on a grid is through nearest neighbor interpolation [163], and has been previously used for 3D reconstruction of B-mode ultrasound images from 2D slices [164]. The algorithm presented here provides an order of magnitude better mean squared reconstruction error and higher signal-to-noise ratio than nearest neighbor interpolation, as demonstrated through simulations and data acquired from a tissue-mimicking (TM) phantom experiment.

7.2 Markov Random Field Model

A Markov random field (MRF) generalizes the one-dimensional Markov property to higher dimensions. This is useful, for instance, when the random process is a function of spatial co-ordinates instead of time. Let $\{X_1, X_2, \ldots, X_t, \ldots\}$ be a sequence of discrete valued random variables forming a Markov chain. Then the joint conditional density

function of a variable conditioned on the past satisfies the following property:

$$P(X_t|X_1,\ldots,X_{t-1}) = P(X_t|X_{t-1}).$$

The same idea can be extended to a "random field" indexed by spatial co-ordinates as follows. Consider an infinite lattice of nodes indexed by triplets $-\infty < i,j,k < \infty$. The MRF property can be written as:

$$\begin{aligned} &P(X_{i^*,j^*,k^*}|\{X_{i,j,k} : i \neq i^*, j \neq j^*, k \neq k^*\}) \\ &\quad = P(X_{i^*,j^*,k^*}|\{X_{i^*\pm1,j^*\pm1,k^*\pm1}\}), \end{aligned}$$

that is, the value at any given node when conditioned on its immediate neighbors is independent of all other node values. These conditions can be analogously defined using continuous density functions (instead of discrete probability mass functions) for random variables defined on a continuous state space. Also, in practice, it is assumed that the grid size is finite $1 \leq i \leq N_y$, $1 \leq j \leq N_x$ and $1 \leq k \leq N_z$ where N_x, N_y and N_z respectively denote the number of grid points along the x, y and z dimensions of the 3D volume.

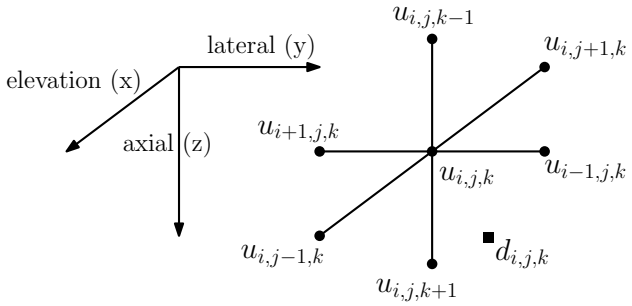


Figure 7.1: A 6-neighborhood of grid nodes is used to calculate the clique potential at each grid point using (7.1). The data point nearest to the central grid point is also used to maintain data fidelity.

In the present case, a 3D lattice Lenz-Ising model is used because the final goal is to reconstruct the measured quantity (shear wave velocities) on a fine grid of points in 3D. A similar idea was used by Gorce *et al.* [165] to model radiofrequency (RF) ultrasound data for spectral processing. Temporarily ignoring the grid points lying on the boundary, a 6-neighborhood of each node as shown in Fig. 7.1 is used to define "clique potentials" as follows [166]:

$$\begin{aligned} V(u_{i,j,k}) = (u_{i,j,k} - d_{i,j,k})^2 \quad &+ \quad \lambda \left[\left(\frac{u_{i,j+1,k} + u_{i,j-1,k} - 2u_{i,j,k}}{\Delta x^2} \right)^2 \right. \\ &+ \quad \left(\frac{u_{i+1,j,k} + u_{i-1,j,k} - 2u_{i,j,k}}{\Delta y^2} \right)^2 \\ &+ \quad \left. \left(\frac{u_{i,j,k+1} + u_{i,j,k-1} - 2u_{i,j,k}}{\Delta z^2} \right)^2 \right], \end{aligned} \tag{7.1}$$

where $d_{i,j,k}$ is the value of the data point closest to the node at lattice location (i, j, k) and $\Delta x, \Delta y, \Delta z$ are the grid resolutions along the three axes. The intuition behind using this potential function is as follows. The first term is a data fidelity term that encourages the node values to be close to the values at nearby data points. The next three terms are second order finite differences that approximate the second derivative at each node location to promote smoothness.

Let $\mathbf{u}$ denote an $N_x N_y N_z \times 1$ vector consisting of the individual $u_{i,j,k}$ values. The vectorization step involves storing the value $u_{i,j,k}$ as the $(i + (j-1)N_y + N_y N_x (k-1))^{th}$ entry in $\mathbf{u}$. The "energy function" is defined as the sum of the potentials over all cliques:

$$E(\mathbf{u}) = \sum_{i,j,k} V(u_{i,j,k}).$$

By the Hammersley-Clifford theorem, the joint density function for the state $\mathbf{u}$ of the MRF can be expressed as [167]:

$$p(\mathbf{u}) = \frac{1}{Z} \exp(-E(\mathbf{u}))$$

where Z is a normalization that ensures that p integrates to unity.

7.3 Iterative Reconstruction Algorithm

The goal of a reconstruction algorithm is to estimate the mode of this distribution, i.e., the value of $\mathbf{u}$ which maximizes $p(\mathbf{u})$, or equivalently, minimizes $-\log p(\mathbf{u}) = \log Z + E(\mathbf{u})$. In most situations, it is impossible to calculate Z explicitly because it involves integrating over all possible node values. However, the value of Z is not required for this minimization problem, and it is sufficient to find a minimizer for $E(\mathbf{u})$.

In practice the reconstruction grid size can be quite large. For instance, even in a simple case of a hundred grid points along each of the three axes, the total number of grid points (i.e. the length of the vector $\mathbf{u}$) is 10^6. A computationally tractable iterative reconstruction approach is shown in Fig. 7.2. The key idea is to start from an initial guess for the solution and then gradually refine the guess such that the energy function value decreases at each iteration. In order to handle large grid sizes, it is important that the updates should be easy to compute, both in terms of memory storage requirement and computational complexity.

Consider the following greedy algorithm [168, Ch. 17] which updates the value at each node myopically, by using values from the previous iteration for each neighbor. Let $\mathbf{u}^{(l)}$ denote the value of the vector at iteration l, for $l \geq 0$. The (i, j, k)th element

of this vector $u_{i,j,k}^{(l+1)}$, is obtained by minimizing the potential function in (7.1) at a fixed coordinate as follows:

$$u_{i,j,k}^{(l+1)} = \arg\min_{u} \quad (u - d_{i,j,k})^2 + \lambda \left[\left(\frac{u_{i,j+1,k}^{(l)} + u_{i,j-1,k}^{(l)} - 2u}{\Delta x^2} \right)^2 + \left(\frac{u_{i+1,j,k}^{(l)} + u_{i-1,j,k}^{(l)} - 2u}{\Delta y^2} \right)^2 + \left(\frac{u_{i,j,k+1}^{(l)} + u_{i,j,k-1}^{(l)} - 2u}{\Delta z^2} \right)^2 \right].$$

This can be solved in closed form by setting the derivative of the function on the right with respect to u to zero:

$$u_{i,j,k}^{(l+1)} = \frac{1}{1 + 4\lambda \left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4} \right)} \left[d_{i,j,k} + \frac{2\lambda}{\Delta x^4} (u_{i,j+1,k}^{(l)} + u_{i,j-1,k}^{(l)}) + \frac{2\lambda}{\Delta y^4} (u_{i+1,j,k}^{(l)} + u_{i-1,j,k}^{(l)}) + \frac{2\lambda}{\Delta z^4} (u_{i,j,k+1}^{(l)} + u_{i,j,k-1}^{(l)}) \right].$$

This is similar to the iterated conditional modes (ICM) technique described by Besag [169]. However, note that the update is only a linear combination of the components of $\mathbf{u}^{(l)}$. Therefore, unlike the original ICM algorithm where each pixel is changed individually, the complete vector can be updated in a single step using the following matrix formulation:

$$\mathbf{u}^{(l+1)} = \frac{1}{1 + 4\lambda \left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4} \right)} \mathbf{d} + \mathbf{A}\mathbf{u}^{(l)} \qquad [7.2]$$

where $\mathbf{d}$ denotes is a vector of $d_{i,j,k}$ values obtained via nearest neighbor interpolation [170, 171]. The matrix $\mathbf{A} = [a_{mn}]$ is defined as:

$$a_{mn} = \begin{cases} \frac{2\lambda/\Delta x^4}{1+4\lambda\left(\frac{1}{\Delta x^4}+\frac{1}{\Delta y^4}+\frac{1}{\Delta z^4}\right)} & \text{if } n = m \pm 1 \\ \frac{2\lambda/\Delta y^4}{1+4\lambda\left(\frac{1}{\Delta x^4}+\frac{1}{\Delta y^4}+\frac{1}{\Delta z^4}\right)} & \text{if } n = m \pm N_y \\ \frac{2\lambda/\Delta z^4}{1+4\lambda\left(\frac{1}{\Delta x^4}+\frac{1}{\Delta y^4}+\frac{1}{\Delta z^4}\right)} & \text{if } n = m \pm N_y N_x \\ 0 & \text{otherwise.} \end{cases} \quad [7.3]$$

where boundary cases are handled by only considering valid indices $1 \leq m, n \leq N_x N_y N_z$. It is worth noting the computational complexity of the update equation (7.2). The matrix $\mathbf{A}$ is $N_x N_y N_z \times N_x N_y N_z$ and the matrix-vector product would require $O(N_x^2 N_y^2 N_z^2)$ multiplications and additions. This can be prohibitive for large grid sizes with millions of grid points. However, note that $\mathbf{A}$ is very sparse and has at most six non-zero entries per row as seen from (7.3). Therefore, each element in $\mathbf{A}\mathbf{u}^{(l)}$ can be calculated in constant time, resulting in an overall complexity of $O(N_x N_y N_z)$.

The following theorem asserts that the sequence of iterates generated in (7.2) must eventually converge to a limit point. This is proved using a contraction mapping argument in the Appendix.

Theorem 7.1 The sequence of iterates $\{\mathbf{u}^{(l)}\}$ generated by the algorithm in Fig. 7.2 has a limit point.

Proof: See Section C.8. □

Input: **d**: vector of nearest neighbor interpolation
A: update matrix defined in (7.3)
$\Delta x, \Delta y, \Delta z$: grid resolutions
λ: smoothing parameter
TOL: tolerance for difference between iterates
Output: **u**: final 3D reconstruction

$l \leftarrow 0$
$\mathbf{u}^{(0)} \leftarrow \mathbf{d}$
repeat
  $\mathbf{u}^{(l+1)} \leftarrow \frac{1}{1+4\lambda\left(\frac{1}{\Delta x^4}+\frac{1}{\Delta y^4}+\frac{1}{\Delta z^4}\right)}\mathbf{d} + \mathbf{A}\mathbf{u}^{(l)}$
  $l \leftarrow l+1$
until $||\mathbf{u}^{(l+1)} - \mathbf{u}^{(l)}|| <$ TOL.
$\mathbf{u} \leftarrow \mathbf{u}^{(l)}$

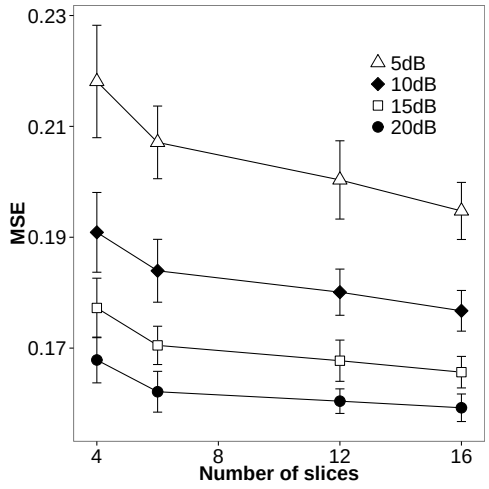
Figure 7.2: Iterative reconstruction algorithm which starts with an initial guess equal to the nearest neighbor interpolation and refines it using a sparse matrix update equation. The stopping criterion used here examines the progress from one iteration to the next and stops if the norm is smaller than a user defined tolerance. Other stopping criteria (such as a fixing the number of iterations ahead of time) may also be used.

7.4 Simulated Ellipsoid Data

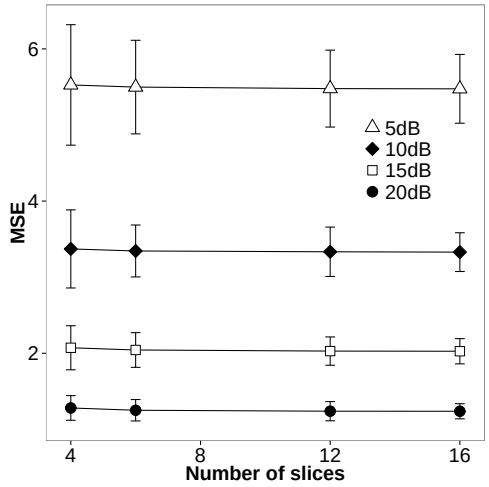
The mean squared error (MSE) performance of the 3D reconstruction algorithm is evaluated using simulated ellipsoidal inclusion data. The "ground truth" SWV values are defined in an ellipsoidal region as follows:

$$f(x,y,z) = \begin{cases} 4 & \text{if } \frac{x^2}{1^2} + \frac{y^2}{1^2} + \frac{(z-2.25)^2}{1.5^2} \leq 1 \\ 1 & \text{otherwise.} \end{cases}$$

These numbers are chosen to closely match the dimensions (in cm) and SWV values (in m/s) of a TM phantom used in Section 7.5. The sheaf acquisition pattern is mimicked by sampling this ellipsoid using a cylindrical co-ordinates grid with 4, 6, 12 and 16 image planes. Each image plane is sampled over a 100×100 grid, 4 cm wide and 4.5 cm high. Simulated i.i.d Gaussian noise is added to this ground truth data with various noise



(a) MRF Algorithm



(b) Nearest neighbor interpolation

Figure 7.3: Simulated mean squared error values from the ellipsoidal inclusion data with different processing parameters. The mean squared error was calculated with respect the ground truth model and the average and standard deviation was calculated over 100 independent realizations with added Gaussian noise. The reconstruction error from a simple nearest neighbor interpolation approach is also shown for comparison.

levels ranging from 5 dB to 20 dB with respect to the inclusion. A 3D reconstruction is generated using the MRF algorithm listed in Fig. 7.2 with $\lambda = 0.01$, and a grid of size $100 \times 100 \times 100$ in a parallelepiped with lateral and elevational dimensions of 4 cm

each and depth of 4.5 cm, so that $\Delta x = \Delta y = 0.04$ cm and $\Delta z = 0.045$ cm. The mean and standard deviation of the reconstruction MSE is calculated over 100 independent realizations of the noisy data. For comparison, the same data is also reconstructed using nearest neighbor interpolation. As seen in Fig. 7.3, in all cases the MRF algorithm provides lower MSE than nearest neighbor interpolation.

7.5 Tissue Mimicking Phantom Experiment

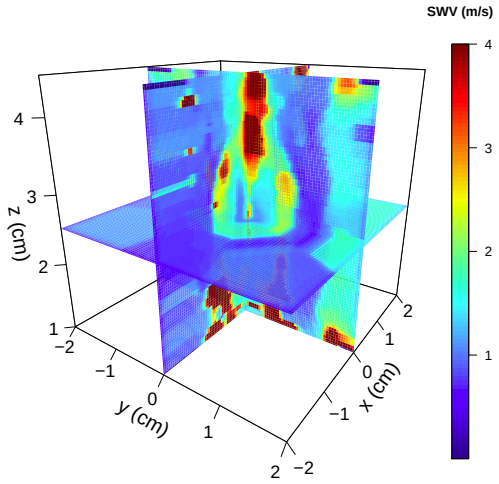
7.5.1 Experimental Setup

An oil-in-gelatin based TM phantom [172] consisting of a stiff ellipsoidal inclusion embedded in a softer background material was used for acquiring EVE data. The phantom consisted of a needle firmly glued to the inclusion to simulate the needle in an ablation procedure. A shear wave pulse is set up in the TM phantom by vibrating the needle using a piezoelectric actuator. The actuator is driven by a controller (Physik Instrumente, Germany) that is synchronized with the ultrasound imaging system (Ultrasonix SonixTOUCH, Canada). A half-sinusoid pulse of amplitude 100 microns and width of 20 ms is used to generate the shear wave.

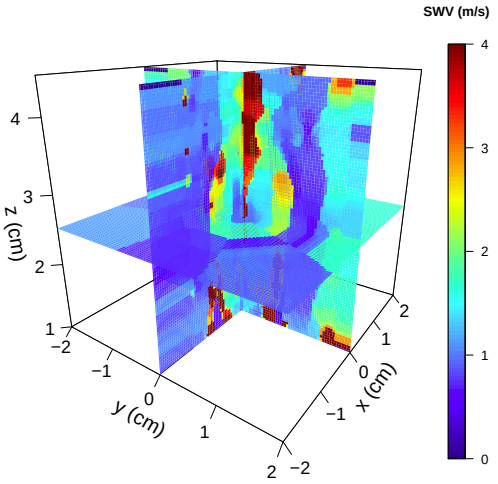
Table 7.1: Shear wave velocity estimates

Algorithm	ROI	4 slices	6 slices	12 slices	16 slices
MRF	bkg	1.06 ± 0.18	1.00 ± 0.15	1.05 ± 0.16	1.05 ± 0.19
	inc	2.03 ± 0.71	1.99 ± 0.67	1.95 ± 0.65	1.97 ± 0.65
NNB	bkg	1.06 ± 0.19	1.00 ± 0.17	1.05 ± 0.19	1.05 ± 0.22
	inc	2.01 ± 0.79	1.98 ± 0.75	1.92 ± 0.76	1.93 ± 0.77

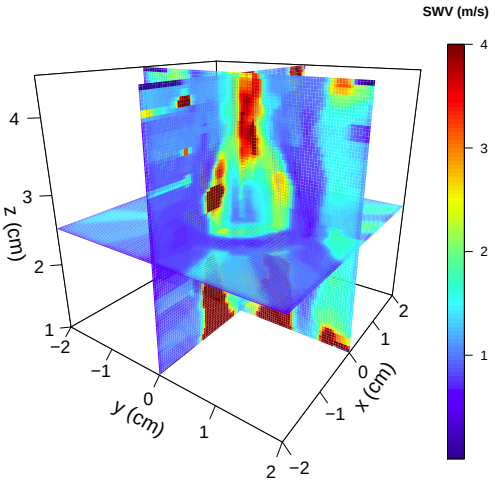
Mean and standard deviation of shear wave velocites (in m/s) obtained from five independent datasets are shown here. The MRF reconstruction has a lower variance in all cases. (MRF=Markov random field algorithm, NNB=nearest neighbors interpolation, bkg= background, inc=inclusion).



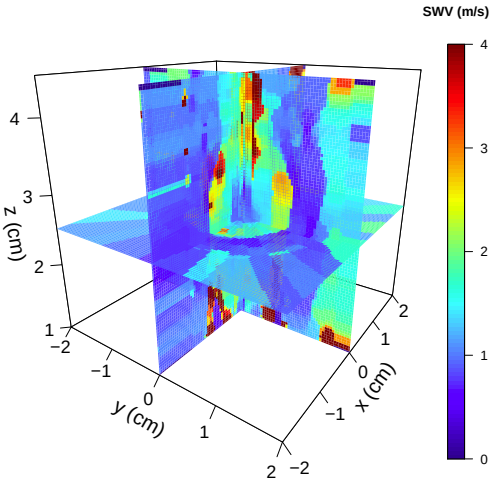
(a) MRF 4 slices



(b) NNB 4 slices



(c) MRF 16 slices



(d) NNB 16 slices

Figure 7.4: 3D reconstructions using 4 slices in (a), (b) and 16 slices in (c), (d) using the MRF reconstruction algorithm in (a), (c) and nearest neighbor interpolation (NNB) in (b), (d). NNB reconstructions appear more blocky as compared to the MRF method, and have poorer signal-to-noise ratio.

The ultrasound system is operated in research mode which allows the use of external triggering and a custom scan sequence. A conventional focussed transmit acquisition

method used to generate B-mode scans does not provide sufficiently high frame rates to track a shear wave. Therefore a phase locked mechanism described in [7] is used to assemble apparent high frame rate RF echo data frames. In this method, multiple pulse vibrations are applied to the needle and only a narrow vertical band of RF data in the image plane is acquired after each vibration. This rests on a reasonable assumption that every pulse has almost identical amplitude and width and there is sufficient delay to allow perturbations from the previous vibration to decay to zero. The 128 element linear array transducer is operated at a center frequency of 5 MHz with the RF data sampled at 40 MHz.

Table 7.2: Contrast and contrast-to-noise ratios

	Algorithm	4 slices	6 slices	12 slices	16 slices
C	MRF	5.70	5.97	5.43	5.53
	NNB	5.56	5.89	5.30	5.33
CNR	MRF	4.04	4.81	3.67	3.27
	NNB	2.08	2.81	1.45	1.03

Values of contrast (C) and contrast-to-noise ratios (CNR) in dB between the inclusion and the background. All numbers are in dB. The Markov random field algorithm provides slightly better C than nearest neighbor interpolation and also provides higher CNR. (MRF=Markov random field algorithm, NNB=nearest neighbor interpolation).

Table 7.3: Signal to noise ratios

ROI	Algorithm	4 slices	6 slices	12 slices	16 slices
bkg	MRF	22.65	23.14	21.72	17.78
	NNB	21.74	21.87	20.24	16.21
inc	MRF	11.14	11.34	10.72	10.47
	NNB	9.03	9.21	8.51	8.22

Signal to noise ratios (SNR) in dB calculated from parallelepiped shaped ROIs in the background and inclusion. The MRF reconstruction has a higher SNR in all cases. (MRF=Markov random field algorithm, NNB=nearest neighbors interpolation, bkg= background, inc=inclusion).

7.5.2 Shear Wave Velocity Reconstruction

The method used for reconstructing individual image planes is identical to the one described in [162]. It is assumed that the shear wave pulse travels laterally in the image plane with the needle acting as a line source and particle displacements are purely axial (aligned with the direction of the ultrasound beam lines). A 1D crosscorrelation-based displacement estimation algorithm with axial window length of 2 mm and 1.5 mm overlap is used to determine particle displacements for each pair of consecutive frames. The resulting frame-to-frame displacement movie can be used to localize the shear wave pulse in both space and time. Due to the lateral propagation assumption, the shear wave trajectory can be traced along lines at constant depth radiating away from the needle. The time of peak displacement at each pixel is used as an estimate for the arrival time of the shear wave pulse with sub-frame-number resolution using a 5-point parabolic fit around the peak.

7.5.3 3D Reconstruction Results

SWV measurements are averaged over five independent datasets obtained on the same phantom to average out any mis-registration errors from manual placement of the transducer probe at different angles [162]. The SWV values from the two reconstruction methods are shown in Table 7.1. Although both methods give similar mean estimates, the MRF algorithm has a lower standard deviation.

Various image quality statistics [144, 138] are calculated to compare the reconstruction quality from the MRF algorithm and the NNB interpolation method. Parallelepiped shaped regions of interest (ROI) of size $5\times5\times10$ mm^3 are used to calculate mean and standard deviations of SWVs in the background and inclusion. Signal to noise ratio

(SNR) is calculated using $\text{SNR}_{inc} = 20\log_{10}\frac{\mu_{inc}}{\sigma_{inc}}$ where the subscript "inc" denotes inclusion. A similar formula is used for the background ("bkg"). Contrast is calculated using $\text{C} = 20\log_{10}\frac{\mu_{inc}}{\mu_{bkg}}$ and contrast-to-noise ratio (CNR) using $\text{CNR} = 20\log_{10}\frac{\mu_{inc}-\mu_{bkg}}{\sqrt{\sigma_{inc}^2+\sigma_{bkg}^2}}$.

Table 7.2 shows C and CNR values for the two methods with different numbers of image slices used for 3D reconstruction. The MRF method provides slightly higher contrast, and also outperforms NNB interpolation in the CNR values by 2 dB. Moreover, its SNR is better by 1 dB in the background region and by 2 dB in the inclusion. These improvements in CNR and SNR can be traced back to the fact that the SWV estimates provided by the MRF algorithm have lower standard deviations because the clique potential model in (7.1) implicitly applies a low pass noise filter by constraining the derivative at each grid point.

Sample 3D reconstructions from the two methods in the extreme cases of 4 and 16 image slices are shown in Fig. 7.4. Reconstructions from NNB interpolation appear blocky, because it fills all the missing grid values with the SWV value from the nearest data point resulting in a "piecewise constant" appearance.

7.6 Discussion

This chapter presented a 3D reconstruction algorithm based on Markov random fields for electrode vibration shear wave elastography. The model-based reconstruction algorithm provides better image quality metrics than standard nearest neighbor interpolation, with improved SNR and CNR. Although the results presented here show SWV data, the reconstruction algorithm is quite general and can be applied to other quantitative measurements such as strain, temperature or shear modulus. Moreover, it can also handle ill-posed inverse problems where the grid size is much larger than the number

of measurements and the measurements may be scattered at arbitrary locations in a 3D volume.

8

Conclusion and Future Work

> The lyf so short, the craft so long to lerne.
>
> Chaucer, *Parlement of Foules*

8.1 Summary of Contributions

A variety of model-based image reconstruction algorithms using tools from statistical signal processing and optimization were developed as part of this dissertation. In Chapter 3, a method for high frame rate data acquisition using plane wave insonification was described, and a prototype was implemented on a research ultrasound scanner. Compared to the traditional focussed transmit acquisition method, plane wave imaging provides lower depth of penetration and signal-to-noise ratios due to the absence of a "narrow pencil-like acoustic beam." The results from this chapter showed that with a few angular plane wave insonifications together with delay-sum beamforming and compounding can be used to track a shear wave pulse in electrode vibration elastography.

Chapter 4 developed a stochastic model for noisy arrival time data which can be used for tracking a shear wave pulse from particle displacement movies. A hidden Markov model framework was presented which encoded the underlying shear wave velocity

(SWV) values as hidden system states that exhibit a "jump behavior." Two different maximum a posteriori estimation algorithms were developed to reconstruct two dimensional (2D) SWV images. Additionally, a brute force optimization based approach was also presented that treated the arrival time information as a piecewise linear function with unknown number and locations of changepoints. These algorithms provide different desirable theoretical properties, while providing good boundary delineation and image quality metrics.

A natural way for extending 2D imaging to three dimensions (3D) using a sheaf acquisition was presented in Chapter 5. Reconstruction quality was gauged using contrast and contrast-to-noise ratio measurements and changes in quality from using increasing number of planes in the sheaf were quantified. The highest contrast of 5 dB was seen between the stiffest and softest regions of the phantom. It was shown that under certain idealizing assumptions on the true shape of the ablation, good reconstruction quality while maintaining fast processing rate can be obtained with as few as 6 imaging planes suggesting that the method is suited for parsimonious data acquisitions with very few sparsely chosen imaging planes.

A model based 3D reconstruction algorithm using the Matérn kernel was developed in Chapter 6. This algorithm rests on the assumption that the transition region between the stiff (ablated) and soft (healthy) tissue is not abrupt, but may be described by a smooth function with a degree of smoothness that can be learned from data. This idea was formalized using reproducing kernel Hilbert space theory, and a smoothing algorithm that automatically estimates the kernel parameters from data was presented. Validation is performed on data acquired from a tissue mimicking phantom. Volume estimates were found to be within 20% of the true value.

Chapter 7 extended the sheaf reconstruction method from Chapter 5 to generate complete 3D visualizations of SWV without resorting to an approximate intermediate step of reconstructing transverse C-planes. The noisy data was modeled using a Markov random field, and a computationally tractable reconstruction algorithm that can handle grids with millions of points was developed. Results from simulated ellipsoidal inclusion data showed that this algorithm outperforms standard nearest neighbor interpolation by an order of magnitude in mean squared reconstruction error. Results from phantom experiments showed that it also provides a higher contrast-to-noise ratio by almost 2 dB and better SNR in the stiff inclusion by over 2 dB.

8.2 Future Work

Bayesian models are quite versatile because they can be used for incorporating prior information that can potentially be used to improve boundary visualization. For instance if additional information such as temperature distribution or strain maps are available, they can be used as priors in a constrained image reconstruction framework. The intuition is that high temperature regions are probably completely ablated and hence should be correlated with the presence of stiffer tissue with low strain values and higher SWVs. In a clinical setting, additional information is often available as pre-procedural CT or MRI scans. This opens the possibility of including SWV images as part of a multi-modality system which can co-register different imaging modalities. This may be used to improve procedure planning and real-time control of percutaneous ablation procedures [173, 174].

The angular plane wave insonification method developed in Chapter 3 was implemented with multiple needle vibrations, where each angle was acquired with a frame rate of around 2 kHz after each vibration. This implementation can be improved by

changing the ultrasound pulse sequence so that the plane waves are immediately fired by looping through the list of angles [111, Fig.13] and compounded using a "slow time" frame-to-frame moving average filter. This method will require only one needle vibration and a higher frame rate by a factor equal to the number of angles used. (For instance, if the acquisition involves 10 insonification angles, the base frame rate must be increased to 20 kHz so that an effective frame rate of 20/10 = 2 kHz is still maintained.) A phased array transducer can be used to obtain greater imaging depth and larger field of view with a sector scan. High frame rate imaging will require development of beamforming algorithms optimized for the phased array geometry. The possibility of improvements in frame rate and image quality through diverging beams [175] and coded excitations [176] should also be explored.

The 3D reconstruction algorithms developed in this dissertation use SWV values estimated from 2D image slices. The next step should involve full wave inversion with a 3D displacement field. This may be achieved through high frame rate volume acquisitions and tracking wave motion laterally and in elevation [142]. The tradeoff between 3D volume rate and image quality should be studied further through simulations and experiments with plane wave imaging sequences with a matrix array transducer.

Although research efforts have not yet conclusively proven the utility of viscoelastic parameters for discriminating between pathological conditions in liver tissue [177, 178], there has been increasing interest in development of methods for measuring the frequency dependence of SWV [179, 180, 181]. These methods make certain assumptions about the viscoelastic model for tissue (such as the Kelvin-Voigt model) and attempt to estimate the complex shear modulus by measuring shear wave and amplitude attenuation as a function of distance from the source and also the wave velocity. The electrode vibration technique used throughout this dissertation can also be used for estimating

viscoelastic properties of tissue, by using continuous vibration at a known frequency and estimating wave velocity as a function of frequency (similar to the concept of dispersion vibrometry [61]). Alternatively, a pulse vibration method can also be used and the deformation in the pulse shape as it travels through the viscoelastic medium can be estimated, in addition to following the peak using the time-to-peak method.

Future work should also explore the clinical applicability of the data acquisition and image reconstruction methods developed in this dissertation. This will require development of a protocol for acquiring data from ablations created *in vivo* in animal models and subsequently through patient studies. Design of a mechanical apparatus for orienting the imaging planes close to the needle should be considered. The sheaf acquisition technique requires accurate knowledge of the angle between different imaging planes, hence it will be advantageous to design an assembly that holds the needle in place and allows the transducer to rotate around the needle through a known angle. A 3-axis positioning system may be needed in free hand elastography applications.

LIST OF REFERENCES

[1] K. E. Maturen, A. P. Wasnik, J. E. Bailey, E. G. Higgins, and J. M. Rubin, "Posterior acoustic enhancement in hepatocellular carcinoma," *J. Ultrasound Med.*, vol. 30, pp. 495–499, Apr. 2011.

[2] C. Cha, J. F. T. Lee, J. Gurney, B. Markhardt, T. Warner, F. Kelcz, and D. Mahvi, "CT versus sonography for monitoring radiofrequency ablation in a porcine liver," *Am. J. Roentgenol.*, vol. 175, pp. 705–711, Sep. 2000.

[3] M. G. van Vledder, M. S. Torbenson, T. M. Pawlik, E. M. Boctor, U. M. Hamper, K. Olino, and M. A. Choti, "The effect of steatosis on echogenicity of colorectal liver metastases on intraoperative ultrasonography," *Arch. Surg.*, vol. 145, pp. 661–667, Jul. 2010.

[4] L. Gao, K. Parker, R. Lerner, and S. Levinson, "Imaging of the elastic properties of tissue—a review," *Ultrasound Med. Biol.*, vol. 22, pp. 959–977, May 1996.

[5] T. Varghese, J. Zagzebski, and F. L. Jr., "Elastographic imaging of thermal lesions in the liver in vivo following radiofrequency ablation: preliminary results," *Ultrasound Med. Biol.*, vol. 28, pp. 1467–1473, Nov.-Dec. 2002.

[6] S. Bharat and T. Varghese, "Radiofrequency electrode vibration-induced shear wave imaging for tissue modulus estimation: A simulation study," *J. Acoust. Soc. Am.*, vol. 128, pp. 1582–1585, Oct. 2010.

[7] R. DeWall, T. Varghese, and E. Madsen, "Shear wave velocity imaging using transient electrode perturbation: Phantom and ex vivo validation," *IEEE Trans. Med. Imag.*, vol. 30, pp. 666–678, Mar. 2011.

[8] A. H. Henni, C. Schmitt, and G. Cloutier, "Three-dimensional transient and harmonic shear-wave scattering by a soft cylinder for dynamic vascular elastography," *J. Acoust. Soc. Am.*, vol. 124, pp. 2394–2405, Oct. 2008.

[9] J. McLaughlin and D. Renzi, "Shear wave speed recovery in transient elastography and supersonic imaging using propagating fronts," *Inv. Prob.*, vol. 22, pp. 681–706, Mar. 2006.

[10] J. Ferlay, H. Shin, F. Bray, D. Forman, C. Mathers, and D. Parkin, "Globocan 2008 v2.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 10 [internet]," 2010. Available from: http://globocan.iarc.fr, accessed on Sep. 11, 2013.

[11] S. Bharat, U. Techavipoo, M. Z. Kiss, W. Liu, and T. Varghese, "Monitoring stiffness changes in lesions after radiofrequency ablation at different temperatures and durations of ablation," *Ultrasound Med. Biol.*, vol. 31, no. 3, pp. 415–422, 2005.

[12] M. Wang, B. Byram, M. Palmeri, N. Rouze, and K. Nightingale, "Imaging transverse isotropic properties of muscle by monitoring acoustic radiation force induced shear waves using a 2-D matrix ultrasound array," *IEEE Trans. Med. Imag.*, vol. 32, pp. 1671–1684, Sep. 2013.

[13] T. J. Hall, A. Oberai, P. E. Barbone, A. M. Sommer, N. H. Gokhale, S. Goenezen, J. Jiang, *et al.*, "Elastic nonlinearity imaging," in *Proc. of IEEE Intl. Conf. of Eng. in Med. and Biol. Soc. 2009*, pp. 1967–1970, 2009.

[14] M. Z. Kiss, T. Varghese, and T. J. Hall, "Viscoelastic characterization of in vitro canine tissue," *Phy. Med. Biol.*, vol. 49, pp. 4207–4218, Sep. 2004.

[15] S. F. Altekruse, K. A. McGlynn, and M. E. Reichman, "Hepatocellular carcinoma incidence, mortality, and survival trends in the united states from 1975 to 2005," *J. Clinical Oncology*, vol. 27, no. 9, pp. 1485–1491, 2009.

[16] J. Ferlay, H. Shin, F. Bray, D. Forman, and C. Parkin, "Globocan 2012, estimated cancer incidence, mortality and preval in 2012 [internet]," 2012. accessed on April 29, 2013.

[17] M. Maluccio and A. Covey, "Recent progress in understanding, diagnosing, and treating hepatocellular carcinoma," *CA: A Cancer Journal for Clinicians*, vol. 62, no. 6, pp. 394–399, 2012.

[18] S. Liu, K. W. Chan, B. Wang, and L. Qiao, "Fibrolamellar hepatocellular carcinoma," *Amer. J. Gastroenterology*, vol. 104, no. 10, pp. 2617–2624, 2009.

[19] S. R. Hamilton, L. A. Aaltonen, I. A. for Research on Cancer, W. H. Organization, *et al.*, *Pathology and genetics of tumours of the digestive system*, vol. 48. IARC press Lyon, 2000.

[20] L. M. Kulik and A. Chokechanachaisakul, "Evaluation and management of hepatocellular carcinoma," *Clinics in Liver Disease*, vol. 19, no. 1, pp. 23–43, 2015. Consultations in Liver Disease.

[21] F. Ferri, "Hepatocellular carcinoma: Treatment," in *Ferriś Clinical Advisor 2015*, Elsevier, 2015.

[22] F. Pons, M. Varela, and J. M. Llovet, "Staging systems in hepatocellular carcinoma," *J. of Intl. Hepato-Pancreato-Biliary Asso.*, vol. 7, no. 1, pp. 35–41, 2005.

[23] E. Loveman, J. Jones, A. J. Clegg, J. Picot, J. Colquitt, D. Mendes, D. J. Breen, E. Moore, S. George, G. Poston, *et al.*, "The clinical effectiveness and cost-effectiveness of ablative therapies in the management of liver metastases: systematic review and economic evaluation," *Health Tech. Assesment*, vol. 18, no. 7, 2014.

[24] S. Weis, A. Franke, J. Mössner, J. C. Jakobsen, and K. Schoppmeyer, "Radiofrequency (thermal) ablation versus no intervention or other interventions for hepatocellular carcinoma," *The Cochrane Library*, 2013.

[25] A. Shen, H. Zhang, C. Tang, Y. Chen, Y. Wang, C. Zhang, and Z. Wu, "A systematic review of radiofrequency ablation versus percutaneous ethanol injection for small hepatocellular carcinoma up to 3cm," *J. Gastroenterology & Hepatology*, vol. 28, no. 5, pp. 793–800, 2013.

[26] C. L. Brace, "Radiofrequency and microwave ablation of the liver, lung, kidney, and bone: what are the differences?," *Current Problems in Diagnostic Radiology*, vol. 38, no. 3, pp. 135–143, 2009.

[27] E. M. Knavel and C. L. Brace, "Tumor ablation: Common modalities and general practices," *Tech. in Vascular & Interventional Radiology*, vol. 16, no. 4, pp. 192–200, 2013. Ablation Update.

[28] T. Shibata, Y. Iimuro, Y. Yamamoto, Y. Maetani, F. Ametani, K. Itoh, and J. Konishi, "Small hepatocellular carcinoma: Comparison of radio-frequency ablation and percutaneous microwave coagulation therapy," *Radiology*, vol. 223, no. 2, pp. 331–337, 2002.

[29] M.-D. Lu, H.-X. Xu, X.-Y. Xie, X.-Y. Yin, J.-W. Chen, M. Kuang, Z.-F. Xu, G.-J. Liu, and Y.-L. Zheng, "Percutaneous microwave and radiofrequency ablation for hepatocellular carcinoma: a retrospective comparative study," *J. Gastroenterology*, vol. 40, no. 11, pp. 1054–1060, 2005.

[30] K. A. Simo, S. E. Sereika, K. N. Newton, and D. A. Gerber, "Laparoscopic-assisted microwave ablation for hepatocellular carcinoma: Safety and efficacy in comparison with radiofrequency ablation," *J. Surgical Oncology*, vol. 104, no. 7, pp. 822–829, 2011.

[31] T. J. Ziemlewicz, J. L. Hinshaw, M. G. Lubner, C. L. Brace, M. L. Alexander, P. Agarwal, and F. T. Lee, "Percutaneous microwave ablation of hepatocellular carcinoma with a gas-cooled system: Initial clinical results with 107 tumors," *J. Vascular and Interventional Radiology*, vol. 26, no. 1, pp. 62–68, 2015.

[32] J. B. Li, J. P. Xu, Y. Chen, and Y. Zhang, "Method and apparatus for needle visualization enhancement in ultrasound images," Dec. 2014. US Patent 20,140,362,114.

[33] G. E. Healthcare, "GE LOGIQe needle recognition." http://www3.gehealthcare.in/en/products/categories/ultrasound/logiq/logiq_e. Accessed: May 19, 2015.

[34] S. Ultrasound, "Sonosite advanced needle visualization." http://www.sonosite.com/needleguide. Accessed: May 19, 2015.

[35] van Vledder MG, T. MS, P. TM, and et al, "The effect of steatosis on echogenicity of colorectal liver metastases on intraoperative ultrasonography," *Archives of Surgery*, vol. 145, no. 7, pp. 661–667, 2010.

[36] N. Rubert and T. Varghese, "Mean scatterer spacing estimation in normal and thermally coagulated ex vivo bovine liver," *Ultrason. Imag.*, vol. 36, no. 2, pp. 79–97, 2014.

[37] T. Varghese, U. Techavipoo, J. A. Zagzebski, and F. T. Lee, "Impact of gas bubbles generated during interstitial ablation on elastographic depiction of in vitro thermal lesions," *Journal of ultrasound in medicine*, vol. 23, no. 4, pp. 535–544, 2004.

[38] T. Wu, J. P. Felmlee, J. F. Greenleaf, S. J. Riederer, and R. L. Ehman, "Assessment of thermal tissue ablation with mr elastography," *Magnetic Resonance in Medicine*, vol. 45, no. 1, pp. 80–87, 2001.

[39] J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: a quantitative method for imaging the elasticity of biological tissues," *Ultrason. Imag.*, vol. 13, no. 2, pp. 111–134, 1991.

[40] I. Cespedes, J. Ophir, H. Ponnekanti, and N. Maklad, "Elastography: elasticity imaging using ultrasound with application to muscle and breast in vivo," *Ultrason. Imag.*, vol. 15, no. 2, pp. 73–88, 1993.

[41] R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," *Science*, vol. 269, no. 5232, pp. 1854–1857, 1995.

[42] Y. K. Mariappan, K. J. Glaser, and R. L. Ehman, "Magnetic resonance elastography: A review," *Clinical Anatomy*, vol. 23, no. 5, pp. 497–511, 2010.

[43] S. K. Venkatesh, M. Yin, and R. L. Ehman, "Magnetic resonance elastography of liver: Technique, analysis, and clinical applications," *Journal of Magnetic Resonance Imaging*, vol. 37, no. 3, pp. 544–555, 2013.

[44] S. Bharat, T. Varghese, E. L. Madsen, and J. A. Zagzebski, "Radio-frequency ablation electrode displacement elastography: a phantom study," *Med. Phy.*, vol. 35, no. 6, pp. 2432–2442, 2008.

[45] S. Bharat, T. G. Fisher, T. Varghese, T. J. Hall, J. Jiang, E. L. Madsen, J. A. Zagzebski, and F. T. L. Jr., "Three-dimensional electrode displacement elastography using the siemens C7F2 foursight four-dimensional ultrasound transducer," *Ultrasound Med. Biol.*, vol. 34, no. 8, pp. 1307 – 1316, 2008.

[46] N. Rubert, S. Bharat, R. DeWall, A. Andreano, C. Brace, J. Jiang, L. Sampson, and T. Varghese, "Electrode displacement strain imaging of thermally-ablated liver tissue in an in vivo animal model," *Med. Phy.*, vol. 37, no. 3, pp. 1075–1082, 2010.

[47] O. Kolokythas, T. Gauthier, A. T. Fernandez, H. Xie, B. A. Timm, C. Cuevas, M. K. Dighe, L. M. Mitsumori, M. F. Bruce, D. A. Herzka, G. K. Goswami, R. T. Andrews, K. M. Oas, T. J. Dubinsky, and B. H. Warren, "Ultrasound-based elastography: a novel approach to assess radio frequency ablation of liver masses performed with expandable ablation probes: A feasibility study," *J. Ultrasound Med.*, vol. 27, pp. 935–946, Jun. 2008.

[48] J. Jiang, C. Brace, A. Andreano, R. J. DeWall, N. Rubert, T. G. Fisher, T. Varghese, F. Lee Jr, and T. J. Hall, "Ultrasound-based relative elastic modulus imaging for visualizing thermal ablation zones in a porcine model," *Physics Med. Biol.*, vol. 55, no. 8, p. 2281, 2010.

[49] L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear modulus imaging with 2-d transient elastography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, pp. 426–435, Apr. 2002.

[50] L. Sandrin, M. Tanter, J.-L. Gennisson, S. Catheline, and M. Fink, "Shear elasticity probe for soft tissues with 1-d transient elastography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, no. 4, pp. 436–446, 2002.

[51] R. J. DeWall, T. Varghese, and C. L. Brace, "Visualizing ex vivo radiofrequency and microwave ablation zones using electrode vibration elastography," *Med. Phy.*, vol. 39, no. 11, pp. 6692–6700, 2012.

[52] K. R. Nightingale, M. L. Palmeri, R. W. Nightingale, and G. E. Trahey, "On the feasibility of remote palpation using acoustic radiation force," *J. Acoust. Soc. Am.*, vol. 110, no. 1, pp. 625–634, 2001.

[53] K. Nightingale, D. Stutz, R. Bentley, and G. Trahey, "Acoustic radiation force impulse imaging: ex vivo and in vivo demonstration of transient shear wave propagation," in *Proc. IEEE Intl. Symp. on Biomed. Imag., 2002 (ISBI 2002)*, pp. 525–528, Jun. 2002.

[54] M. L. Palmeri, M. H. Wang, J. J. Dahl, K. D. Frinkley, and K. R. Nightingale, "Quantifying hepatic shear modulus in vivo using acoustic radiation force," *Ultrasound Med. Biol.*, vol. 34, pp. 546–558, Apr. 2008.

[55] J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: a new technique for soft tissue elasticity mapping," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 51, no. 4, pp. 396–409, 2004.

[56] L. Ji, J. R. McLaughlin, D. Renzi, and J.-R. Yoon, "Interior elastodynamics inverse problems: shear wave speed reconstruction in transient elastography," *Inv. Prob.*, vol. 19, no. 6, p. S1, 2003.

[57] P. Song, H. Zhao, A. Manduca, M. Urban, J. Greenleaf, and S. Chen, "Comb-push ultrasound shear elastography (cuse): A novel method for two-dimensional shear elasticity imaging of soft tissues," *IEEE Trans. Med. Imag.*, vol. 31, pp. 1821–1832, Sep. 2012.

[58] C. Hazard, Z. Hah, D. Rubens, and K. Parker, "Integration of crawling waves in an ultrasound imaging system. part 1: System and design considerations," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 296–311, 2012.

[59] Z. Hah, C. Hazard, B. Mills, C. Barry, D. Rubens, and K. Parker, "Integration of crawling waves in an ultrasound imaging system. part 2: Signal processing and applications," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 312–323, 2012.

[60] M. Orescanin, M. Qayyum, K. Toohey, and M. Insana, "Dispersion and shear modulus measurements of porcine liver," *Ultrason. Imag.*, vol. 32, pp. 255–266, Oct. 2010.

[61] S. Chen, M. Urban, C. Pislaru, R. Kinnick, Y. Zheng, A. Yao, and J. Greenleaf, "Shearwave dispersion ultrasound vibrometry (sduv) for measuring tissue elasticity and viscosity," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, pp. 55–62, Jan. 2009.

[62] C. Amador, M. Urban, S. Chen, and J. Greenleaf, "Shearwave dispersion ultrasound vibrometry (sduv) on swine kidney," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, pp. 2608–2619, Dec. 2011.

[63] C. T. Barry, B. Mills, Z. Hah, R. A. Mooney, C. K. Ryan, D. J. Rubens, and K. J. Parker, "Shear wave dispersion measures liver steatosis," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 175–182, 2012.

[64] S. McAleavey, M. Menon, and E. Elegbe, "Shear modulus imaging with spatially-modulated ultrasound radiation force," *Ultrason. Imag.*, vol. 31, no. 4, pp. 217–234, 2009.

[65] M. H. Wang, M. L. Palmeri, V. M. Rotemberg, N. C. Rouze, and K. R. Nightingale, "Improving the robustness of time-of-flight based shear wave speed reconstruction methods using RANSAC in human liver in vivo," *Ultrasound Med. Biol.*, vol. 36, pp. 802–813, May 2010.

[66] N. Rouze, M. Wang, M. Palmeri, and K. Nightingale, "Robust estimation of time-of-flight shear wave speed using a Radon sum transformation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, pp. 2662–2670, Dec. 2010.

[67] J. Klein, J. McLaughlin, and D. Renzi, "Improving arrival time identification in transient elastography," *Phys. Med. Biol.*, vol. 57, pp. 2151–2168, Apr. 2012.

[68] H. Zhao, P. Song, M. W. Urban, R. R. Kinnick, M. Yin, J. F. Greenleaf, and S. Chen, "Bias observed in time-of-flight shear wave speed measurements using radiation force of a focused ultrasound beam," *Ultrasound Med. Biol.*, vol. 37, no. 11, pp. 1884–1892, 2011.

[69] K. Nightingale, S. McAleavey, and G. Trahey, "Shear-wave generation using acoustic radiation force: in vivo and ex vivo results,," *Ultrasound Med. Biol.*, vol. 28, pp. 1715–1723, Dec. 2003.

[70] T. Deffieux, J.-L. Gennisson, B. Larrat, M. Fink, and M. Tanter, "The variance of quantitative estimates in shear wave elastography: Theory and experiments," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 59, pp. 2390–2410, Nov. 2012.

[71] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[72] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Am. Stat. Asso.*, vol. 74, no. 368, pp. 829–836, 1979.

[73] M. Ratkovic and K. Eng, "Finding jumps in otherwise smooth curves," *Political Analysis*, vol. 18, pp. 57–77, 2010.

[74] J. Bai and P. Perron, "Computation and analysis of multiple structural change models," *J. Appl. Econometrics*, vol. 18, no. 1, pp. 1–22, 2003.

[75] D. Hudson, "Fitting segmented curves whose join points have to be estimated," *Amer. Stat. Asso.*, vol. 61, no. 316, pp. 1097–1129, 1966.

[76] R. Bellman and R. Roth, "Curve fitting by segmented straight lines," *Amer. Stat. Asso.*, vol. 64, no. 327, pp. 1079–1084, 1969.

[77] A. Gallant and W. Fuller, "Fitting segmented polynomial regression models whose join points have to be estimated,," *J. Amer. Stat. Asso.*, vol. 68, no. 341, pp. 144–147, 1973.

[78] A. Gholami and S. M. Hosseini, "A balanced combination of tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals," *Signal Processing*, vol. 93, no. 7, pp. 1945–1960, 2013.

[79] D. Denison, B. Mallick, and A. Smith, "Automatic Bayesian curve fitting," *J. Royal Stat. Soc. B*, vol. 60, no. 2, pp. 333–350, 1998.

[80] A. Tishler and I. Zang, "A new maximum likelihood algorithm for piecewise regression,," *J. Amer. Stat. Asso.*, vol. 76, pp. 980–987, Dec. 1981.

[81] A. Magnani and S. Boyd, "Convex piecewise-linear fitting,," *Optim Eng.*, vol. 10, pp. 1–17, Mar. 2009.

[82] J. Friedman, "Multivariate adaptive regression splines,," *Ann. of Stats.*, vol. 19, pp. 1–67, Mar. 1991.

[83] E. Kolaczyk and R. Nowak, "Multiscale generalized linear models for nonparametric function estimation,," *Biometrica*, vol. 92, pp. 119–133, Mar. 2005.

[84] A. Saucier and C. Audet, "Construction of sparse signal representations with adaptive multiscale orthogonal bases,," *Signal Processing*, vol. 92, pp. 1446–1457, Jun. 2012.

[85] J. Bai and P. Perron, "Estimating and testing linear models with multiple structural changes," *Econometrica*, vol. 66, pp. 47–78, Jan. 1998.

[86] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions on*, vol. 19, pp. 716–723, Dec. 1974.

[87] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using mcmc with applications to signal segmentation,," *IEEE Trans. Sig. Proc.*, vol. 50, pp. 747–758, Mar. 2002.

[88] P. Fearnhead, "Exact Bayesian curve fitting and signal segmentation,," *IEEE Trans. Sig. Proc.*, vol. 53, pp. 2160–2166, Jun. 2005.

[89] H. J. Kwon, M. J. Kang, J. H. Cho, J. Y. Oh, K. J. Nam, S. Y. Han, and S. W. Lee, "Acoustic radiation force impulse elastography for hepatocellular carcinoma-associated radiofrequency ablation," *World J. Gastroenterol.*, vol. 17, pp. 1874–1878, Apr. 2011.

[90] S. T. Elliott, "Volume ultrasound: the next big thing?," *Br. J. Radiol.*, vol. 81, pp. 8–9, Jan. 2008.

[91] H. Cardinal, J. Gill, and A. Fenster, "Analysis of geometrical distortion and statistical variance in length, area, and volume in a linearly scanned 3-D ultrasound image," *IEEE Trans. Med. Imag.*, vol. 19, pp. 632–651, Jun. 2000.

[92] R. S. José-Estépar, M. Martin-Fernández, P. Caballero-Martinez, C. Alberola-López, and J. Ruiz-Alzola, "A theoretical framework to three-dimensional ultrasound reconstruction from irregularly sampled data," *Ultrasound Med. Biol.*, vol. 29, pp. 255–269, Feb. 2003.

[93] J. E. Lindop, G. M. Treece, A. H. Gee, and R. W. Prager, "3D elastography using freehand ultrasound," *Ultrasound Med. Biol.*, vol. 32, pp. 529–545, Apr. 2006.

[94] H. Rivaz, I. Fleming, L. Assumpcao, G. Fichtinger, U. Hamper, M. Choti, G. Hager, and E. Boctor, "Ablation monitoring with elastography: 2D in-vivo and 3D ex-vivo studies," *Medical Image Computing & Computer-Assisted Intervention–MICCAI 2008*, vol. 11, no. 2, pp. 458–466, 2008.

[95] M.-A. Janvier, G. Soulez, L. Allard, and G. Cloutier, "Validation of 3D reconstructions of a mimicked femoral artery with an ultrasound imaging robotic system," *Med. Phys.*, vol. 37, pp. 3868–3879, Jun. 2010.

[96] S. Bharat, T. G. Fisher, T. Varghese, T. J. Hall, J. Jiang, E. L. Madsen, J. A. Zagzebski, and F. T. Lee Jr, "Three-dimensional electrode displacement elastography using the Siemens C7F2 fourSight four-dimensional ultrasound transducer," *Ultrasound Med. Biol.*, vol. 34, pp. 1307–1316, Aug. 2008.

[97] P. Foroughi, G. Hager, and E. Boctor, "Robust elasticity imaging using external tracker," in *Prod. IEEE Intl. Symp. on Biomed. Imag., 2009 (ISBI 2009)*, pp. 209–212, Jun. 2009.

[98] A. V. Patil, C. D. Garson, and J. A. Hossack, "3D prostate elastography: algorithm, simulations and experiments," *Phys. Med. Biol.*, vol. 52, pp. 3643–3663, Jun. 2007.

[99] L. Zhai, J. Madden, W. C. Foo, V. Mouraviev, T. J. Polascik, M. L. Palmeri, and K. R. Nightingale, "Characterizing stiffness of human prostates using acoustic radiation force," *Ultrason. Imag.*, vol. 32, pp. 201–213, Oct. 2010.

[100] S. H. Lee, J. M. Chang, W. H. Kim, M. S. Bae, N. Cho, A. Yi, H. R. Koo, S. J. Kim, J. Y. Kim, and W. K. Moon, "Differentiation of benign from malignant solid breast masses: comparison of two-dimensional and three-dimensional shear-wave elastography," *Eur. Radiol.*, vol. 23, pp. 1015–1026, Apr. 2013.

[101] D. Dotti, E. Gatti, V. Svelto, A. Ugge, and P. Vidali, "Blood-flow measurements by ultrasound correlation techniques," *Energia Nucleare*, vol. 23, no. 11, pp. 571–575, 1976.

[102] L. Wilson and D. Robinson, "Ultrasonic measurement of small displacements and deformations of tissue," *Ultrasonic imaging*, vol. 4, no. 1, pp. 71–82, 1982.

[103] T. Krouskop, D. Dougherty, F. Vinson, *et al.*, "A pulsed doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue," *J Rehabil Res Dev*, vol. 24, no. 2, pp. 1–8, 1987.

[104] E. Madsen, G. Frank, T. Krouskop, T. Varghese, F. Kallel, and J. Ophir, "Tissue-mimicking oil-in-gelatin dispersions for use in heterogeneous elastography phantoms," *Ultrason. Imag.*, vol. 25, no. 1, pp. 17–38, 2003.

[105] R. Sinkus, J. Lorenzen, D. Schrader, M. Lorenzen, M. Dargatz, and D. Holz, "High-resolution tensor MR elastography for breast tumour detection," *Phys. Med. Biol.*, vol. 45, pp. 1649–1664, Jun. 2000.

[106] M. Tanter, M. Pernot, G. Montaldo, J.-L. Gennisson, E. Bavu, E. Mace, T.-M. Nguyen, M. Couade, and M. Fink, "Real time quantitative elastography using supersonic shear wave imaging," in *Proc. 2010 IEEE Intl. Symp. on Biomed. Imag.: From Nano to Macro*, pp. 276–279, Apr. 2010.

[107] J. A. Jensen, S. I. Nikolov, K. L. Gammelmark, and M. H. Pedersen, "Synthetic aperture ultrasound imaging," *Ultrasonics*, vol. 44, pp. e5–e15, 2006.

[108] J. Lin, J. R. McLaughlin, D. Renzi, and J.-R. Yoon, "Interior elastodynamics inverse problems: shear wave speed reconstruction in transient elastography," *Inv. Prob.*, vol. 19, no. 6, pp. S1–S29, 2003.

[109] J. R. McLaughlin and D. Renzi, "Using level set based inversion of arrival times to recover shear wave speed in transient elastography and supersonic imaging," *Inv. Prob.*, vol. 22, no. 2, pp. 707–725, 2006.

[110] A. Ingle, J. Bucklew, W. Sethares, and T. Varghese, "Slope estimation in noisy piecewise linear functions," *Signal Processing*, vol. 108, pp. 576–588, Mar. 2015.

[111] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 3, pp. 489–506, 2009.

[112] A. Ingle and T. Varghese, "A comparison of model based and direct optimization based filtering algorithms for shearwave velocity reconstruction for electrode vibration elastography," in *Proc. IEEE Intl. Symp. on Biomed. Imag., 2013 (ISBI 2013)*, pp. 760–763, 2013.

[113] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer, 2nd ed., 2006.

[114] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, May 1983.

[115] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Sig. Proc.*, vol. 50, pp. 174–188, Feb. 2002.

[116] W. R. Gilks and C. Berzuini, "Following a moving targetmonte carlo inference for dynamic bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001.

[117] N. Bergman, A. Doucet, and N. Gordon, "Optimal estimation and cramér-rao bounds for partial non-gaussian state space models," *Ann. Inst. Stat. Math.*, vol. 53, no. 1, pp. 97–112, 2001.

[118] B. Banerjee, D. Roy, and R. M. Vasu, "Efficient implementations of a pseudodynamical stochastic filtering strategy for static elastography," *Med. Phy.*, vol. 36, no. 8, pp. 3470–3476, 2009.

[119] T. E. Marchant, A. Skalski, and B. Matuszewski, "Automatic tracking of implanted fiducial markers in cone beam ct projection images," *Med. Phy.*, vol. 39, no. 3, pp. 1322–1334, 2012.

[120] T. Raveendran, S. Gupta, R. M. Vasu, and D. Roy, "A pseudo-time enkf incorporating shape based reconstruction for diffuse optical tomography," *Med. Phy.*, vol. 39, no. 2, pp. 1092–1101, 2012.

[121] B. Byram, G. Trahey, and M. Palmeri, "Bayesian speckle tracking. part I: an implementable perturbation to the likelihood function for ultrasound displacement estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 60, Jan. 2013.

[122] B. Byram, G. E. Trahey, and M. Palmeri, "Bayesian speckle tracking. part II: biased ultrasound displacement estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 60, no. 1, 2013.

[123] A. Smith, A. Doucet, N. de Freitas, and N. Gordon, "Improving regularized particle filters," in *Sequential Monte Carlo methods in practice*, Springer Science & Business Media, 2013.

[124] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Prob. and Appl.*, vol. 14, no. 1, pp. 153–158, 1969.

[125] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Stat. and Comput.*, vol. 10, no. 3, pp. 197–208, 2000.

[126] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition,," *Proceedings of the IEEE*, vol. 77, pp. 257–286, Feb. 1989.

[127] R. Bellman, *Dynamic Programming,*. Mineola, NY:: Dover, 2003.

[128] S. Boyd, L. Vandenberghe, and C. O. Problems, *in: Convex Optimization,*. Cambridge, UK:: Cambridge University Press, 1 ed., 2004.

[129] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm,," *J. Royal Stat. Soc.*, vol. 39, 1977.

[130] W. Qian and D. Titterington, "Estimation of parameters in hidden Markov models,," *Philos Trans. Roy. Soc. London Ser. A*, vol. 337, pp. 407–428, 1991.

[131] G. Archer and D. Titterington, "Parameter estimation for hidden Markov chains," *J. Stat. Planning and Inference*, vol. 108, pp. 365–390, 2002.

[132] C. F. J. Wu, "On the convergence properties of the em algorithm,," *Ann. of Stat.*, vol. 11, pp. 95–103, Mar. 1983.

[133] G. Goodwin and J. Agüero, "Approximate em algorithms for parameter and state estimation in nonlinear stochastic models," in *Proc. 44th IEEE Conference on Decision and Control*, (Seville, Spain, December), pp. 12–15, 2005.

[134] T.-J. Yen, "A majorization-minimization approach to variable selection using spike and slab priors,," *Ann. of Stat.*, vol. 39, pp. 1748–1755, Jun. 2011.

[135] I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures,," *Statistics and Decisions*, vol. Suppl. 1, pp. 205–237, 1984.

[136] R. Yeung and T. B.-A. Algorithms, *Information Theory and Network Coding,*. Springer, 1 ed., 2008. Ch. 9 Sec. 9.3, pp. 222–226.

[137] J. Bucklew, "Two results on asymptotic performance of quantizers,," *IEEE Trans. Info. Theory*, vol. 30, p. 2, Mar. 1984.

[138] T. Varghese and J. Ophir, "An analysis of elastographic contrast to noise ratio," *Ultrasound Med. Biol.*, vol. 24, pp. 915–924, Jul. 1998.

[139] T. Varghese, "Quasi-static ultrasound elastography," *Ultrasound Clinics*, vol. 4, pp. 323–338, Jul. 2009.

[140] E. L. Madsen, M. A. Hobson, H. Shi, T. Varghese, and G. R. Frank, "Stability of heterogeneous elastography phantoms made from oil dispersions in aqueous gels," *Ultrasound Med. Biol.*, vol. 32, no. 2, pp. 261–270, 2006.

[141] R. J. DeWall, T. Varghese, and C. L. Brace, "Quantifying local stiffness variations in radiofrequency ablations with dynamic indentation," *IEEE Trans. Biomed. Eng.*, vol. 59, pp. 728–735, Mar. 2012.

[142] J.-l. Gennisson, J. Provost, T. Deffieux, C. Papadacci, M. Imbault, M. Pernot, M. Tanter, *et al.*, "4-D ultrafast shear-wave imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, no. 6, pp. 1059–1065, 2015.

[143] P. Abolmaesumi, S. Salcudean, W.-H. Zhu, M. Sirouspour, and S. DiMaio, "Image-guided control of a robot for medical ultrasound," *IEEE Trans. Robot. Autom.*, vol. 18, pp. 11–23, Feb. 2002.

[144] T. Varghese and J. Ophir, "A theoretical framework for performance characterization of elastography: the strain filter," *IEEE Trans. Ultrason., Ferroelectr., Freq Control*, vol. 44, pp. 164–172, Jan. 1997.

[145] M. Elkateb Hachemi, S. Calle, and J. P. Remenieras, "Transient displacement induced in shear wave elastography: comparison between analytical results and ultrasound measurements," *Ultrasonics*, vol. 44 Suppl 1, pp. e221–225, Dec. 2006.

[146] G. Wahba, *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1990.

[147] L. Wasserman, *All of Nonparametric Statistics (Springer Texts in Statistics)*. Springer, Oct. 2005.

[148] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics, Springer, 2013.

[149] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, "Real-time regularized ultrasound elastography," *IEEE Trans. Med. Imag.*, vol. 30, pp. 928–945, Apr. 2011.

[150] H. Rivaz, E. Boctor, P. Foroughi, R. Zellars, G. Fichtinger, and G. Hager, "Ultrasound elastography: A dynamic programming approach," *IEEE Trans. Med. Imag.*, vol. 27, pp. 1373–1377, Oct. 2008.

[151] V. Naumova, S. V. Pereverzyev, and S. Sivananthan, "Extrapolation in variable rkhss with application to the blood glucose reading," *Inv. Prob.*, vol. 27, no. 7, pp. 1–13, 2011.

[152] J. Carew, R. Dalal, G. Wahba, and S. Fain, "A nonparametric method for estimation of arterial wall shear stress," in *Proc. of International Soc. of Magnetic Resonance in Med. (ISMRM)*, vol. 12, p. 1924, 2004.

[153] G. Wahba and Y. Wang, "When is the optimal regularization parameter insensitive to the choice of the loss function?," *Commun. in Stat. - Theory and Methods*, vol. 19, no. 5, pp. 1685–1700, 1990.

[154] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *J. Math. Anal. & Appl.*, vol. 33, Jan. 1971.

[155] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999. Properties of Random Fields, ch. 2, sec. 2.7, pp. 31–33.

[156] B. Minasny and A. B. McBratney, "The Matérn function as a general model for soil variograms," *Geoderma*, vol. 128, no. 3-4, pp. 192–207, 2005.

[157] S. Ramani, D. Van De Ville, T. Blu, and M. Unser, "Nonideal sampling and regularization theory," *IEEE Trans. Sig. Proc.*, vol. 56, no. 3, pp. 1055–1070, 2008.

[158] R. S. Strichartz, *A guide to distribution theory and Fourier transforms*. Singapore: World Scientific, 2003.

[159] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions With Formulars, Graphs, and Mathematical Tables*, vol. 55. Dover Publications, 1964.

[160] F. Utreras, "On generalized cross-validation for multivariate smoothing spline functions," *SIAM J. Scientific and Stat. Computing*, vol. 8, no. 4, pp. 630–643, 1987.

[161] D. Yang, M. Converse, D. Mahvi, and J. Webster, "Measurement and analysis of tissue temperature during microwave liver ablation," *IEEE Trans. Biomed. Eng.*, vol. 54, pp. 150–155, Jan. 2007.

[162] A. Ingle and T. Varghese, "Three-dimensional sheaf of ultrasound planes reconstruction (SOUPR) of ablated volumes," *IEEE Trans. Med. Imag.*, vol. 33, pp. 1677–1688, Aug. 2014.

[163] S. Arya and D. M. Mount, "Approximate nearest neighbor queries in fixed dimensions," in *Proc. 4th Annual ACM-SIAM Symp. on Discrete Algorithms*, SODA 1993, (Philadelphia, PA, USA), pp. 271–280, 1993.

[164] J. Blancher, C. Léger, and L. D. Nguyen, "Time-varying, 3-d echocardiography using a fast-rotating probe," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 51, no. 5, pp. 634–639, 2004.

[165] J.-M. Gorce, D. Friboulet, I. Dydenko, J. D'hooge, B. Bijnens, and I. Magnin, "Processing radio frequency ultrasound images: a robust method for local spectral features estimation by a spatially constrained parametric approach," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, pp. 1704–1719, Dec. 2002.

[166] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[167] G. R. Grimmett, "A theorem about random fields," *Bul. of the London Math. Soc.*, vol. 5, no. 1, pp. 81–84, 1973.

[168] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*. McGraw-Hill Higher Education, 2 ed., 2001.

[169] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.

[170] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, pp. 509–517, Sep. 1975.

[171] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J. of the ACM*, vol. 45, no. 6, pp. 891–923, 1998.

[172] E. Madsen, G. Frank, M. Hobson, H. Shi, J. Jiang, T. Varghese, and T. Hall, "Spherical lesion phantoms for testing the performance of elastography systems," *Phy. Med. Biol.*, vol. 50, pp. 5983–5995, Dec. 2005.

[173] L. Crocetti, R. Lencioni, S. DeBeni, T. C. See, C. Della Pina, and C. Bartolozzi, "Targeting liver lesions for radiofrequency ablation: an experimental feasibility study using a ct–us fusion imaging system," *Investigative radiology*, vol. 43, no. 1, pp. 33–39, 2008.

[174] H. Elhawary, S. Oguro, K. Tuncali, P. R. Morrison, S. Tatli, P. B. Shyn, S. G. Silverman, and N. Hata, "Multimodality non-rigid image registration for planning, targeting and monitoring during ct-guided percutaneous liver tumor cryoablation," *Acad. Radiology*, vol. 17, no. 11, pp. 1334–1344, 2010.

[175] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, pp. 102–119, Jan. 2014.

[176] P. Song, M. Urban, A. Manduca, J. Greenleaf, and S. Chen, "Coded excitation plane wave imaging for shear wave motion detection," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, pp. 1356–1372, Jul. 2015.

[177] S. Chen, W. Sanchez, M. R. Callstrom, B. Gorman, J. T. Lewis, S. O. Sanderson, J. F. Greenleaf, H. Xie, Y. Shi, M. Pashley, *et al.*, "Assessment of liver viscoelasticity by using shear waves induced by ultrasound radiation force," *Radiology*, vol. 266, no. 3, pp. 964–970, 2013.

[178] X. Chen, Y. Shen, Y. Zheng, H. Lin, Y. Guo, Y. Zhu, X. Zhang, T. Wang, and S. Chen, "Quantification of liver viscoelasticity with acoustic radiation force: a study of hepatic fibrosis in a rat model," *Ultrasound Med. Biol.*, vol. 39, no. 11, pp. 2091–2102, 2013.

[179] M. Selzo and C. Gallippi, "Viscoelastic response (visr) imaging for assessment of viscoelasticity in voigt materials," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 60, pp. 2488–2500, Dec. 2013.

[180] S.-W. Huang, H. Xie, J.-L. Robert, S. Zhou, and V. Shamdasani, "Exact viscoelastic green's functions of the voigt-model-based navier's equation," in *IEEE Ultrason. Symp. 2013 (IUS 2013)*, pp. 352–355, Jul. 2013.

[181] J. Langdon, E. Elegbe, and S. Mcaleavey, "Single tracking location acoustic radiation force impulse viscoelasticity estimation (stl-ve): A method for measuring tissue viscoelastic parameters," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, pp. 1225–1244, Jul. 2015.

[182] T. Deffieux, J.-L. Gennisson, J. Bercoff, and M. Tanter, "On the effects of reflected waves in transient shear wave elastography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, no. 10, pp. 2032–2035, 2011.

[183] R. J. DeWall and T. Varghese, "Improving thermal ablation delineation with electrode vibration elastography using a bidirectional wave propagation assumption," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 59, no. 1, pp. 168–173, 2012.

[184] M. Tanter, J. Bercoff, A. Athanasiou, T. Deffieux, J.-L. Gennisson, G. Montaldo, M. Muller, A. Tardivon, and M. Fink, "Quantitative assessment of breast lesion viscoelasticity: initial clinical results using supersonic shear imaging," *Ultrasound in medicine & biology*, vol. 34, no. 9, pp. 1373–1386, 2008.

[185] W. Rudin, *Principles of mathematical analysis*. New York: McGraw-Hill Book Co., third ed., 1987. International Series in Pure and Applied Mathematics.

[186] F. Marvasti, "Extension of lagrange interpolation to 2-d nonuniform samples in polar coordinates," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 567–568, Apr. 1990.

[187] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.*, vol. 69, pp. 1519–1525, Nov. 1979.

[188] R. J. Marks, "Polar, spiral and generalized sampling and interpolation," in *Advanced topics in Shannon sampling and interpolation theory*, Springer-Verlag, 1993.

[189] E. Margolis and Y. Eldar, "Nonuniform sampling in polar coordinates with applications to computerized tomography," in *Proc. of IEEE Convention of Electr. Electron. Eng. in Israel 2004*, pp. 372–375, Sep. 2004.

[190] C. D. Meyer, "Perron-frobenius theory," in *Matrix Analysis and Applied Linear Algebra*, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.

[191] J. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle,," *Econometrica*, vol. 57, pp. 357–384, Mar. 1989.

[192] R. Garica and P. Perron, "An analysis of the real interest rate under regime shifts," *Review of Econ. and Stat.*, vol. 78, pp. 111–125, Feb. 1996.

[193] J. Bai and P. Perron, "strucchange RealInt dataset." http://svitsrv25.epfl.ch/R-doc/library/strucchange/html/RealInt.html, 2003. Accessed: Jul 1, 2015.

A

Artifacts in Shear Wave Imaging

Shear wave speed estimation from high frame rate ultrasound elastography data is a challenging inverse problem due to the requirement of taking derivatives and reciprocals with noisy data. Some of the sources of artifacts in shear wave imaging are, physical limitations of the ultrasound imaging modality, reconstruction algorithm used for solving the inverse problem, inaccurate assumptions about mechanical properties of tissue and limitations of the method of shear wave generation. A short review of existing literature on sources and mitigation of such artifacts is presented in the following paragraphs.

Two commonly used techniques proposed in literature for localizing shear waves in space and time are crosscorrelation based time of arrival method [109] and the time to peak method [54]. Both methods suffer from artifacts resulting from reflection of shear waves traveling through disparate media. For example, [9] demonstrate the presence of false peaks in the reconstructed shear wave velocity maps, using both synthetic and experimental data. The correlation based wave localization procedure makes use of both phase and amplitude information of the "template" shear wave pulse. Although this results in an algorithm that is more robust to noise, it causes confusion from the wave traveling in the opposite direction after reflection. Even though this reflected wave may have much smaller amplitude than the forward traveling wave, the exact alignment of phase causes a false peak in the correlation values. Reflection artifacts may be mitigated

using frequency domain directional filters on the particle velocity maps as proposed in ultrasound elastography by Deffieux *et al.* [182] and demonstrated experimentally by Song *et al.* [57].

Tracking shear waves in homogeneous media is slightly less challenging than the case of disparate media where the locations and number of interfaces is unknown. Nightingale *et al.* presented two different algorithms for estimation of shear wave velocity in homogeneous media that are robust to outliers: the Radon sum method [66] and RANSAC method [65]. They filter the noisy displacement maps using a "quadratic motion filter" that may be the cause for a bias in the final velocity estimates.

Inclusion boundary blurring can be observed in some shear wave velocity images especially when the spatial frequency of the shear wave is low. This is a physical limitation of the imaging technique due to the fact that the resolution is limited by the wavelength of the shear wave pulse. Moreover, as noted in [9], the boundary blurriness is accompanied by overestimation in the wave velocity estimate in the interior of a stiff inclusion. These artifacts are worse when the inclusion is stiffer because of extreme bending of of the velocity level curves that may be confused as self intersecting loops due to measurement noise. In a later paper, the same authors suggest the use of total variation smoothing to mitigate some of these artifacts.

High velocity artifacts can result above and below the inclusion boundaries if the pure lateral propagation assumption does not hold. This is typically the case in electrode vibration elastography where particle motion at locations deeper than the end of the ablation needle is not purely transverse and may have a lateral component. Such artifacts can be reduced by assuming two orthogonal velocity components and picking the component with better SNR [183]. Wave speed inverse problems are often simplified by

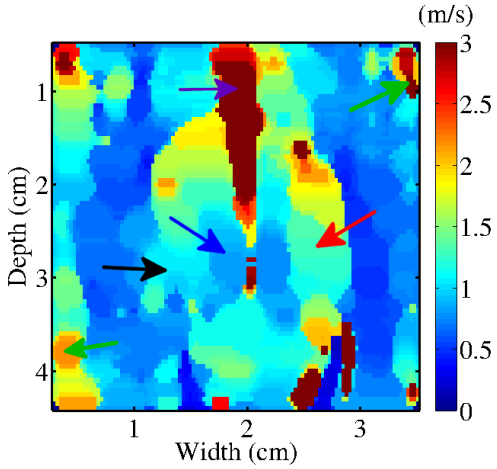


Figure A.1: This figure shows various structures and artifacts present in the SWV image plane. The red arrow points to the stiff ablated region and the black arrow points to the partially ablated region. There is a low velocity artifact near the needle shown by the blue arrow at the center. There are high velocity artifacts close to the edge of the image shown with green arrows in the top right and bottom left corners. The violet arrow near the top of the image plane points to a high velocity artifact due to the needle.

making certain idealizing assumptions about tissues being linearly elastic, incompressible and ignore viscosity. Although these assumptions are not detrimental to phantom experiments, they do cause additional artifacts when testing data from real tissues. In viscoelastic media, shear wave speed is frequency dependent. This is manifested through the change in shape of the shear wave shape as it travels away from the source. A modified search algorithm [184, 67] allows a scaling factor for the wave template to obtain better crosscorrelation values.

Various artifacts can also be seen in the SWV maps and the C-plane reconstructions. These are summarized in Fig. A.1. High velocity artifacts appear above and below the inclusion because the assumption of purely lateral shear wave propagation may not hold those regions. Removal of such artifacts has been studied previously [141] and can be incorporated into the 3D reconstruction algorithm. High velocity artifacts are also seen close to the edge of the image plane due to limited imaging aperture in those areas.

Methods for removal of artifacts due to wavefront distortion and reflection have been discussed previously [9, 67] and have the potential to improve reconstruction quality for the 3D case too. Low velocity artifacts seen in areas adjacent to the needle have been observed in previous EVE studies [7] and require further investigation. The authors conjecture that the shear wave pulse takes some time to accelerate to its maximum speed after the needle is vibrated resulting in perceived low velocity close to the needle source. This artifact may not be a serious hurdle in application because regions immediately adjacent to the needle are certainly ablated; confirming the location of the outer boundary of the ablation is more crucial. Results in a recent paper by Deffieux *et al.* [70] suggest that the uncertainty in measuring higher shear wave velocities is inherently higher. This indicates there is a lower limit to how accurately SWV maps can be reconstructed inside the stiff inclusion.

B

Algorithm Pseudocode

B.1 Particle Filter Algorithm

Input: $\{M_k^i, \omega_k^i\}$: set of particles and weights
Output: $\{M_k^i, \omega_k^i\}$: updated set of particles and weights

procedure RESAMPLE($\{M_k^i, \omega_k^i\}$)
    Initialize cumulative distribution function $c_1 \leftarrow 0$
    for $i = 2 : N_s$ **do**
        $c_i \leftarrow c_{i-1} + w_k^i$
    end for
    Draw u_1 uniformly randomly from $[0, 1/N_s]$
    for $i = 1 : N_s$ **do**
        $u_i \leftarrow u_1 + (i-1)/N_s$
        while $u_i > c_j$ **do**
            $j \leftarrow j + 1$
        end while
        $M_k^i \leftarrow M_k^j$
        $\omega_k^i \leftarrow 1/N_s$
    end for
end procedure

Figure B.1: Resampling algorithm used within the particle filter algorithm shown in Fig. B.3

Input: $\{M_k^i, \omega_k^i\}_{i=1}^{N_s}$: set of particles and weights
Output: $\{\omega_k^i\}_{i=1}^{N_s}$: set of backward smoothed weights

1: **procedure** BACKWARDSMOOTH($\{M_k^i, \omega_k^i\}_{i=1}^{N_s}$)
2: **for** $k = N - 1 : 1$ **do**
3: **for** $i = 1 : N_s$ **do**
4: Recalculate $\omega_k^i \leftarrow \sum_{j=1}^{N_s} \omega_{k+1}^j \frac{\omega_k^i p(M_{k+1}^j | M_k^i)}{\sum_{l-1}^{N_s} \omega_k^l p(M_{k+1}^j | M_k^l)}$
5: **end for**
6: **end for**
7: **end procedure**

Figure B.2: Backward smoothing routine used in the particle filter algorithm shown in Fig. B.3

Input: $X_{1:N}$: noisy data vector of length N
N_s: number of particles to use
N_T: threshold for resampling
σ^2: Gaussian noise variance
p: probability of staying in the same slope value
$[LB, UB]$: lower and upper bounds for slope values
Output: $M_{1:N}$: optimal state sequence

procedure ParticleSmoother($\{X_k\}_{k=1}^N, N_s, N_T, p, \sigma^2, [UB, LB]$)
  for $k = 1 : N$ **do**
    for $i = 1 : N_s$ **do**
      Randomly draw $M_k^i \sim p(M_k^i | M_{k-1}^i)$ Eq. (4.4)
      Update weight ω_k^i using Eq. (4.9) for each $i = 1 : N_s$
    end for
    Normalize all weights $\omega_k^i \leftarrow \omega_k^i / \sum_{i=1}^{N_s} \omega_k^i$
    Calculate effective sample size $N_{eff} \leftarrow \frac{1}{\sum_{i=1}^{N_s} (\omega_k^i)^2}$.
    if $N_{eff} < N_T$ **then**
      Covariance matrix $S_k \leftarrow cov(M_k^i, \omega_k^i)$ and Cholesky factorization $D_k D_k^T = S_k$
      $\{M_k^i, \omega_k^i\}_{i=1}^{N_s} \leftarrow$ Resample($\{M_k^i, \omega_k^i\}_{i=1}^{N_s}$)
      for $j = 1 : N_s$ **do**
        Draw $\epsilon^i \sim$ Epanechnikov Kernel
        $M_k^i \leftarrow M_k^i + D_k \epsilon^i$
      end for
    end if
  end for
  $\{\omega_k^i\}_{i=1}^{N_s} \leftarrow$ BackwardSmooth($\{M_k^i, \omega_k^i\}_{i=1}^{N_s}$)
  for $k = 1 : N$ **do**
    Calculate optimal state $M_k \leftarrow \sum_{i=1}^{N_s} \omega_k^i M_k^i$ by Eq. (4.10).
  end for
end procedure

Figure B.3: Details of the particle filter algorithm in pseudocode adapted from Algorithm 6 in the paper by Arulampalam *et al.*[115] and Section V of Doucet *et al.*[125].

B.2 MAPSlope-FastTrellis Algorithms

Input: $x_{1:N}$: noisy data vector of length N
M: number of slope quantization levels
σ^2: Gaussian noise variance
p: probability of staying in the same slope value
Output: $s^*_{1:N}$: MAP slope sequence
L^*: log-likelihood for MAP slope sequence

procedure FastTrellis($x_{1:N}, M, \sigma^2, p$)
  $I(0,0) \leftarrow 0$
  for $n = 1$ to N **do**
    for $k = 0$ to $(M-1)n$ **do**
      if $n == 1$ **then**
        $I(1,k) \leftarrow -\frac{1}{2\sigma^2}\left(x_1 - \frac{k}{M-1}\right)^2 + \log\frac{1}{M}$
        $\mathbf{\Pi}(1,k) \leftarrow [0,k]$.
      else
        $I(n,k) \leftarrow \max_j[I(n-1,j) + \text{branch reward of } (n-1,j) \text{ to } (n,k)]$
        $\widehat{j} \leftarrow \arg\max_j[I(n-1,j) + \text{branch reward of } (n-1,j) \text{ to } (n,k)]$
        $\mathbf{\Pi}(n,k) \leftarrow$ Append$(\mathbf{\Pi}(n-1,\widehat{j}),k)$
      end if
    end for
  end for
  $L^* \leftarrow \max_k I(N,k)$
  $s^*_{1:N} \leftarrow \mathbf{\Pi}(N, \arg\max_k I(N,k))$
  return $(s^*_{1:N}, L^*)$
end procedure

Figure B.4: Fast dynamic program for searching an optimal route through a linearly growing trellis (like the one shown in Fig. 4.3).

Input: $x_{1:N}$: noisy data vector of length N
M: number of slope quantization levels
σ^2: initial guess for the Gaussian noise variance
p: guess probability of staying in the same slope value
τ: threshold for likelihood value convergence test
MAXIT: maximum number of iteration
Output: $s^*_{1:N}$: MAP slope sequence
σ^{*2}: estimated noise variance
p^*: estimated value of p

procedure MAPSLOPE($x_{1:N}, M, \sigma^2, p, \tau$, MAXIT)
  loop $\leftarrow 1$
  $L \leftarrow -\infty$
  CONVERGED $\leftarrow$ *False*
  repeat
    if using FASTTRELLIS **then**
      $(s^*_{1:N}, L^*) \leftarrow$FASTTRELLIS$(x_{1:N}, M, \sigma^2, p)$
    else if using another optimization routine **then**
      $s^*_{1:N} \leftarrow \arg\max_{s_{1:N}} \log \mathsf{p}(x_{1:N}, s_{1:N}|\sigma^2, p)$
      $L^* \leftarrow \log \mathsf{p}(x_{1:N}, s^*_{1:N}|\sigma^2, p)$
    end if
    $\sigma^{*2} \leftarrow \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \sum_{j=1}^{i} s^*_j\right)^2$
    $p^* \leftarrow \frac{1}{N-1}\sum_{n=1}^{N-1}\mathbf{1}_{\{0\}}(s^*_{n+1} - s^*_n)$
    loop $\leftarrow$ loop $+ 1$
    if $(|L^* - L| < \tau) \vee (\text{loop} > \text{MAXIT})$ **then**
      CONVERGED $\leftarrow$ *True*
    end if
    $p \leftarrow p^*$, $\sigma^2 \leftarrow \sigma^{*2}$, $L \leftarrow L^*$
  until $\neg$CONVERGED
  return $(s^*_{1:N}, \sigma^{*2}, p^*)$
end procedure

Figure B.5: MAPSLOPE algorithm for maximum a posteriori slope estimation in piecewise linear functions with alternating maximization parameter estimation.

C

Proofs and Derivations

C.1 Derivation of EM Iterations

The complete data likelihood function is given by

$$\begin{aligned}
\mathsf{p}(x_{1:N}, s_{1:N}|\theta) &= \prod_{i=1}^{N} \mathsf{p}(x_i|s_{1:N},\theta) \prod_{k=2}^{N} \mathsf{p}(s_k|s_{k-1},\theta) \\
&= \prod_i f\left(x_i; \sum_{j=1}^{i} s_j, \sigma^2\right) \prod_k \left[p\, \mathbf{1}_{\{0\}}(s_k - s_{k-1}) + \frac{1-p}{M-1}\, \mathbf{1}^c_{\{0\}}(s_k - s_{k-1})\right]
\end{aligned}$$

where $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ is the univariate Gaussian probability density function parametrized by its mean and variance.

Next, the auxiliary function for EM iterations is constructed as follows:

$$\begin{aligned}
Q(\theta,\theta') &= \sum_{s_{1:N}\in\mathcal{S}^N} \left[\log \mathsf{p}(x_{1:N}, s_{1:N}|\theta)\right] \mathsf{p}(x_{1:N}, s_{1:N}|\theta') \\
&= \sum_{s_{1:N}\in\mathcal{S}^N} \sum_i \log f\left(x_i, \sum_{j=1}^{i} s_j, \sigma^2\right) \mathsf{p}(x_{1:N}, s_{1:N}|\theta') \\
&\quad + \sum_{s_{1:N}\in\mathcal{S}^N} \sum_k \log\left[p\, \mathbf{1}_{\{0\}}(s_k - s_{k-1}) + \frac{1-p}{M-1}\, \mathbf{1}^c_{\{0\}}(s_k - s_{k-1})\right] \\
&\qquad\qquad \cdot \mathsf{p}(x_{1:N}, s_{1:N}|\theta') \qquad\qquad [C.1] \\
&=: T_1 + T_2
\end{aligned}$$

where $\mathcal{S}^N$ is the set of all valid slope value sequences of length N, and the primes denote old values of the parameters from the previous iteration.

The two terms can now be independently optimized due to decoupling of parameters σ^2 and p.

Estimating σ^2 Setting the gradient of T_1 with respect to σ^2 to zero yields

$$\sigma^2 = \frac{\sum_{n=1}^{N} \sum_{s_{1:n} \in \mathcal{S}^n} \left(x_n - \sum_{j=1}^{n} s_j \right)^2 \mathsf{p}(x_{1:N}, s_{1:n} | \theta')}{N \mathsf{p}(x_{1:N} | \theta')}$$

The constraint $\sigma^2 > 0$ is automatically met since all the terms in the expression on the right hand side are positive.

Estimating p Setting the gradient of T_2 with respect to p to zero yields

$$p = \frac{\sum_{k=2}^{N} \sum_{i=j} \mathsf{p}(x_{1:N}, S_{k-1} = i, S_k = j | \theta')}{(N-1) \mathsf{p}(x_{1:N} | \theta')}.$$

The constraint $0 < p < 1$ is automatically met since all the quantities in the expression on the right hand side of this equation are positive and

$$\sum_{i=j} \mathsf{p}(x_{1:N}, S_{k-1} = i, S_k = j | \theta') \leq \sum_{i,j} \mathsf{p}(x_{1:N}, S_{k-1} = i, S_k = j | \theta') = \mathsf{p}(x_{1:N} | \theta').$$

C.2 Convergence of MAPSlope-FastTrellis

C.2.1 Proof of Theorem 4.2

It suffices to prove the two inequalities for $\log \mathsf{p}(x_{1:N}, s_{1:N}|\theta)$ because log is monotonic increasing. The first inequality follows from the fact that FastTrellis solves the maximization problem

$$s^*_{1:N} = \arg\max_{s_{1:N}} \log \mathsf{p}(x_{1:N}, s_{1:N}|\theta).$$

The second inequality can be obtained maximizing $\log \mathsf{p}(x_{1:N}, s^*_{1:N}|\theta)$ as a function of θ. Setting the derivative with respect to σ to zero yields,

$$-\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}\left(x_i - \sum_{j=1}^{i} s^*_j\right)^2 = 0$$

which gives

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \sum_{j=1}^{i} s^*_j\right)^2.$$

Also note that the second derivative with respect to σ is negative implying that this is in fact a maximum.

Next, setting the derivative with respect to p to zero yields,

$$\sum_{k:\, s_k = s_{k-1}, 2\le k\le N} \frac{1}{p} = \sum_{k:\, s_k \neq s_{k-1}, 2\le k\le N} \frac{1}{1-p}$$

which gives

$$p = \frac{1}{N-1}\sum_{k=2}^{N} \mathbf{1}_{\{0\}}(s^*_k - s^*_{k-1}).$$

Again, note that the second derivative with respect to p is negative which implies that this is a maximizer. □

C.2.2 Proof of Corollary 4.3

Assuming there exists $\delta > 0$ such that $\sigma \geq \delta$, the following upper bound is obtained:

$$\log \mathsf{p}(x_{1:N}, s_{1:N}|\theta) \leq -\frac{N}{2}\log(2\pi\delta^2) + (N-1)\max\left(\log p, \log\frac{1-p}{M-1}\right)$$

for all $s_{1:N}$ and θ. Since every bounded non-decreasing sequence has a limit point by the monotone convergence theorem [185, Theorem 3.14], it follows from Theorem 4.2 that the alternating maximization iterations must converge. □

C.3 Least-squares Minimization Problem in (5.1)

We want to minimize $||\mathbf{Ax}-\mathbf{b}||^2 + \eta||\mathbf{Bx}||^2$ by choice of $\mathbf{x}$.

$$\begin{aligned} ||\mathbf{Ax}-\mathbf{b}||^2 &= (\mathbf{Ax}-\mathbf{b})^T(\mathbf{Ax}-\mathbf{b}) + \eta(\mathbf{Bx})^T(\mathbf{Bx}) \\ &= \mathbf{x}^T(\mathbf{A}^T\mathbf{A}+\eta\mathbf{B}^T\mathbf{B})\mathbf{x} - 2\mathbf{b}^T\mathbf{Av} + \mathbf{b}^\mathbf{T}\mathbf{b}. \end{aligned}$$

Setting the derivative with respect to $\mathbf{x}$ to zero, yields:

$$2(\mathbf{A}^T\mathbf{A}+\eta\mathbf{B}^T\mathbf{B})\mathbf{x} - 2\mathbf{A}^T\mathbf{b} = 0$$

implying that

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A}+\eta\mathbf{B}^T\mathbf{B})^{-1}\mathbf{A}^T\mathbf{b}.$$

C.4 A Sampling Theorem for the Sheaf Geometry

For the sheaf geometry, it is natural to express the shear wave velocity values as a function of cylindrical coordinates (ρ, φ, z) with the needle coinciding with the z-axis. While the sampling in ρ and z directions are controlled by the sampling resolution of the ultrasound elastogram, the sampling in the angular φ direction is controlled by the number of planes in the sheaf chosen for scanning. The conditions under which exact reconstruction is possible (at least in theory) can be stated using Shannon's sampling theorem extended to 3D in cylindrical coordinates. Such results have been previously reported for 2D [186, 187, 188].

Theorem C.1 (Sheaf Sampling Theorem) Let $x : \mathbb{R}_+ \times [0, 2\pi) \times \mathbb{R} \to \mathbb{R}$ be a function defined using a cylindrical co-ordinate system on three dimensional euclidean space. The function values are sampled at equally spaced points over a cylindrical volume using an equiangular sheaf of planes. The cylinder has diameter $2R$ and depth D. Let $x(\rho, \varphi, z)$ be the value of this function at a point (ρ, φ, z), $\rho \geq 0$, $\varphi \in [0, 2\pi)$ and $-\infty < z < \infty$. Let $f_{(\varphi,z)}(\rho) = x(\rho, \varphi, z)$ for φ, z fixed and similarly define $g_{(\rho,z)}(\varphi)$ and $h_{(\rho,\varphi)}(z)$. Also, let $L^2(A)$ denote the class of functions defined on the domain A that are square integrable.

(A1) For every (φ, z), $f_{(\varphi,z)}(\rho) \in L^2$, is bandlimited to W_ρ and the radial sampling rate $F_\rho > 2W_\rho$.

(A2) For every (ρ, φ), $h_{(\rho,\varphi)}(z) \in L^2$, is bandlimited to B_z and the height sampling rate $F_z > 2W_z$.

(A3) For every (ρ, z), $g_{(\rho,z)}(\varphi) \in L^2$ and the Fourier series of its 2π-periodic extension is limited to $\pm N$ terms.

If (A1)–(A3) hold then x can be reconstructed from its samples using an interpolation formula:

$$\begin{aligned} x(\rho,\varphi,z) &= \sum_{k=-\infty}^{\infty}\sum_{r=-N}^{N}\sum_{l=-\infty}^{\infty} x\left(\frac{k}{F_\rho},\frac{2\pi r}{2N+1},\frac{l}{F_z}\right)\operatorname{sinc}(F_\rho\rho-k) \\ &\quad\times\frac{\sin(N+\frac{1}{2})(\phi-\frac{2\pi r}{2N+1})}{\sin(\frac{\varphi}{2}-\frac{2\pi r}{2(2N+1)})}\operatorname{sinc}(F_z z-l). \end{aligned} \tag{C.2}$$

A detailed proof adapted from ideas in [189] is shown below.

Proof: For fixed ρ and z, if $g_{(\rho,z)}(\varphi)$ is in $L^2[0,2\pi)$, its 2π-periodic extension can be expanded as a Fourier series:

$$g_{(\rho,z)}(\varphi)=\sum_{n=-\infty}^{\infty} d_n(\rho,z)e^{jn\varphi}$$

where $j=\sqrt{-1}$. By assumption (A3), this sum can be truncated to $|n|\le N$ so that

$$g_{(\rho,z)}(\varphi)=\sum_{n=-N}^{N} d_n(\rho,z)e^{jn\varphi}. \tag{C.3}$$

Define a $2N+1$ dimensional complex valued vector $\mathbf{d}=(d_{-N},\cdots,d_N)$. Each component of this vector is a function of (ρ,z). Let $\omega=e^{j2\pi/(2N+1)}$ be the complex $(2N+1)^{th}$ root of unity and define $\mathbf{e}_r=(\omega^{-rN},\cdots,\omega^{rN})$. The set of vectors $\{\mathbf{e}_r\}_{r=-N}^{N}$ forms an orthogonal basis for $\mathbb{R}^{2N+1}$ and hence $\mathbf{d}$ can be expanded as:

$$\begin{aligned} \mathbf{d} &= \frac{1}{2N+1}\sum_{r=-N}^{N}\langle d,\mathbf{e}_r\rangle\mathbf{e}_r \\ &= \frac{1}{2N+1}\sum_{r=-N}^{N}\sum_{l=-N}^{N} d_l\omega^{-lr}\mathbf{e}_r \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2N+1} \sum_{r=-N}^{N} \sum_{l=-N}^{N} d_l e^{-j\frac{2\pi lr}{2N+1}} \mathbf{e}_r \\
&= \frac{1}{2N+1} \sum_{r=-N}^{N} g_{(\rho,z)}\left(\frac{-2\pi r}{2N+1}\right) \mathbf{e}_r.
\end{aligned}
$$

Here $\langle \cdot, \cdot \rangle$ denotes the inner product. The n^{th} component of $\mathbf{d}$ can be expressed as:

$$
d_n(\rho, z) = \frac{1}{2N+1} \sum_{r=-N}^{N} g_{(\rho,z)}\left(\frac{-2\pi r}{2N+1}\right) e^{j\frac{2\pi n}{2N+1}r}.
$$

Substituting this into (C.3) yields,

$$
\begin{aligned}
g_{(\rho,z)}(\varphi) &= \sum_{n=N}^{N} \frac{1}{2N+1} \sum_{r=-N}^{N} g_{(\rho,z)}\left(\frac{-2\pi r}{2N+1}\right) e^{j\frac{2\pi n}{2N+1}r} e^{jn\varphi} \\
&= \frac{1}{2N+1} \sum_{r=-N}^{N} x\left(\rho, \frac{-2\pi r}{2N+1}, z\right) \frac{\sin(N+\frac{1}{2})(\varphi - \frac{2\pi r}{2N+1})}{\sin\left(\frac{\varphi}{2} + \frac{2\pi r}{2(2N+1)}\right)}. \qquad \text{[C.4]}
\end{aligned}
$$

For each fixed $r \in \{-N, \cdots, N\}$, $x\left(\rho, \frac{-2\pi r}{2N+1}, z\right)$ is bandlimited both as a function of ρ and z by assumptions (A1) and (A2). Let X be the 2D Fourier transform of x as a function of ρ and z. Then,

$$
x\left(\rho, \frac{-2\pi r}{2N+1}, z\right) = \int_{-W_\rho}^{W_\rho} \int_{-W_z}^{W_z} X(\eta, \xi)\, e^{j2\pi(\rho\eta + z\xi)}\, d\eta\, d\xi.
$$

Since $F_\rho > 2W_\rho$ and $F_z > 2W_z$, $X(\cdot, \cdot) \in L^2([-\frac{F_\rho}{2}, \frac{F_\rho}{2}] \times [-\frac{F_z}{2}, \frac{F_z}{2}])$, and it can be expanded as a 2D Fourier series:

$$
X(\eta, \xi) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} c_{kl}\, e^{-j2\pi\left(\frac{k\eta}{F_\rho} + \frac{l\xi}{F_z}\right)}
$$

where $c_{kl} = \frac{1}{F_\rho F_z} x\left(\frac{k}{F_\rho}, \frac{-2\pi r}{2N+1}, \frac{l}{F_z}\right)$. Define the partial sums

$$X_m(\eta,\xi) = \sum_{k=-m}^{m}\sum_{l=-m}^{m} c_{kl}\, e^{-j2\pi\left(\frac{k\eta}{F_\rho}+\frac{l\xi}{F_z}\right)}$$

and note that $\lim_{m\to\infty} X_m = X$ in L^2. Then,

$$\begin{aligned}
x\left(\rho, \frac{-2\pi r}{2N+1}, z\right) &= \langle X(\eta,\xi), e^{-j2\pi(\rho\eta+z\xi)}\rangle \\
&= \lim_{m\to\infty}\langle X_m(\eta,\xi), e^{-j2\pi(\rho\eta+z\xi)}\rangle \\
&= \lim_{m\to\infty}\int_{-\frac{F_\rho}{2}}^{\frac{F_\rho}{2}}\int_{-\frac{F_z}{2}}^{\frac{F_z}{2}}\sum_{k=-m}^{m}\sum_{l=-m}^{m} c_{kl} e^{-j2\pi\left(\frac{k\eta}{F_\rho}+\frac{l\xi}{F_z}\right)} e^{-j2\pi(\rho\eta+z\xi)} d\eta d\xi \\
&= \sum_k\sum_l \frac{1}{F_\rho F_z} x\left(\frac{k}{F_\rho}, \frac{-2\pi r}{2N+1}, \frac{l}{F_z}\right)\int_{-\frac{F_\rho}{2}}^{\frac{F_\rho}{2}}\int_{-\frac{F_z}{2}}^{\frac{F_z}{2}} e^{j2\pi(\eta(\rho-\frac{k}{F_\rho})+\xi(z-\frac{l}{F_z}))} d\eta d\xi \\
&= \sum_k\sum_l x\left(\frac{k}{F_\rho}, \frac{-2\pi r}{2N+1}, \frac{l}{F_z}\right)\mathrm{sinc}(F_\rho\rho-k)\mathrm{sinc}(F_z z-l) \qquad \text{[C.5]}
\end{aligned}$$

Substituting (C.5) in (C.4) gives the interpolation formula (C.2) which completes the proof.

□

Remark C.2 (Application) In the special case of data acquired in an ultrasound scan, the sampling rate conditions can be interpreted in terms of elastogram resolution. Let L be the lateral width of the transducer, D be the imaging depth and K be the number of planes imaged in an equiangular sheaf. Then conditions (A1)–(A3) require each plane to be sampled over a grid of size at least $2W_\rho L \times 2W_z D$ and number of angular slices $K \geq N$.

Remark C.3 The following theorem is stated in more generality and applies to measurements such as strain, shear wave velocity, stiffness modulus, or TTP that may be measured in 3D over a subset of a sheaf.

C.5 Proof of Theorem 6.2

Define the space $\mathcal{H}_0 = \{f \in \mathcal{H}_R : f = \sum_{i=1}^{n} a_i R(t, t_i)\}$. Clearly $\mathcal{H}_0$ is a subspace of $\mathcal{H}_R$ and one can write $\mathcal{H}_R = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_1$ consists of all the functions that are orthogonal to the functions in the subspace $\mathcal{H}_0$. So, any $g \in \mathcal{H}_R$ can be represented as $g = g_0 + g_1$ such that $||g_0 + g_1||^2_{\mathcal{H}_R} = ||g_0||^2_{\mathcal{H}_R} + ||g_1||^2_{\mathcal{H}_R}$. Moreover $g(t_i) = g_0(t_i)$ because point evaluation at t_i is equivalent to an inner product operation with the reproducing kernel $R(\cdot, t_i) \in \mathcal{H}_0$. This implies that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(u_i - g(t_i))^2 + \lambda||g||^2_{\mathcal{H}_R} &= \frac{1}{n}\sum_{i=1}^{n}(u_i - g_0(t_i))^2 + \lambda(||g_0||^2_{\mathcal{H}_R} + ||g_1||^2_{\mathcal{H}_R}) \\ &\geq \frac{1}{n}\sum_{i=1}^{n}(u_i - g_0(t_i))^2 + \lambda||g_0||^2_{\mathcal{H}_R} \end{aligned}$$

so that the solution must lie in $\mathcal{H}_0$. The solution is unique because the objective function in (6.1) is convex and coercive. □

C.6 Proof of Theorem 6.4

By definition of the gamma function,

$$\Gamma(\nu + 1) = \int_0^\infty u^\nu e^{-u}\, du$$

where, substituting $u = t(1/r^2 + \omega_1^2 + \omega_2^2)$ and rearrangement yields:

$$(1/r^2 + \omega_1^2 + \omega_2^2)^{-\nu-1} = \frac{1}{\Gamma(\nu+1)} \int_0^\infty t^\nu e^{-t(1/r^2+\omega_1^2+\omega_2^2)}\, dt. \tag{C.6}$$

The inverse 2D Fourier transform is given by

$$\begin{aligned}
& \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \frac{e^{j(\omega_1 x_1 + \omega_2 x_2)}}{(1/r^2 + \omega_1^2 + \omega_2^2)^{\nu+1}}\, d\omega_1\, d\omega_2 && \text{[C.7]} \\
= & \frac{1}{(2\pi)^2 \Gamma(\nu+1)} \int_{t=0}^\infty t^\nu e^{-\frac{t}{r^2}} \int_{\mathbb{R}^2} e^{-t(\omega_1^2+\omega_2^2)} e^{j(\omega_1 x_1 + \omega_2 x_2)}\, d\omega_1\, d\omega_2\, dt && \text{[C.8]} \\
= & \frac{1}{(2\pi)^2 \Gamma(\nu+1)} \int_{t=0}^\infty t^\nu e^{-\frac{t}{r^2} - \frac{x_1^2+x_2^2}{4t}} \int_{\mathbb{R}^2} e^{-t((\omega_1 - j\frac{x_1}{2t})^2 + (\omega_2 - j\frac{x_2}{2t})^2)} d\omega_1 d\omega_2 dt && \text{[C.9]} \\
= & \frac{1}{4\pi\Gamma(\nu+1)} \int_{t=0}^\infty t^{\nu-1} e^{-\frac{t}{r^2} - \frac{x_1^2+x_2^2}{4t}}\, dt && \text{[C.10]} \\
= & \frac{1}{2\pi\Gamma(\nu+1)} \int_0^\infty 2 \left(\frac{\sqrt{x_1^2+x_2^2}}{2/r} \right)^\nu \cosh(u\nu) e^{-\frac{1}{r}\sqrt{x_1^2+x_2^2}\cosh(u)}\, du && \text{[C.11]} \\
= & \frac{1}{2^{\nu+1}\pi\Gamma(\nu+1) r^{-2\nu}} \left(\frac{\sqrt{x_1^2+x_2^2}}{r} \right)^\nu K_\nu \left(\frac{\sqrt{x_1^2+x_2^2}}{r} \right) && \text{[C.12]} \\
= & \frac{\Gamma(\nu+\frac{1}{2})}{4\pi^{3/2}\Gamma(\nu+1)} R(||(x_1,x_2) - (0,0)||) && \text{[C.13]}
\end{aligned}$$

where (C.8) is obtained by substituting (C.6) in (C.7); (C.10) follows from the fact that the inner integral is the integral of a 2D Gaussian function which evaluates to π/t; (C.11) is obtained by substituting $t = \frac{1}{2} r e^{-u} \sqrt{x_1^2 + x_2^2}$ and using the fact that $\cosh(u)$ is an even function; and (C.12) follows from the definition of the modified Bessel function of the second kind [159, Sec. 9.6.24, pp. 376]. Finally, (C.13) follows from the definition of $R(\cdot)$ in (6.2).

C.7 Proof of Corollary 6.6

Fix any $(\nu, r) \in \mathcal{S}_1 \times \mathcal{S}_2$. Then from Theorem 6.5 as $n \uparrow \infty$, $I_{\nu,r}(n) \downarrow 1$ where the dependence of the inefficiency function on ν and r is explicitly shown. Since $R_{\nu,r}$ is the reproducing kernel for $\mathcal{W}^{1/r}_{(\nu+1),2}$, solving the optimization problem (6.1) produces a function $f \in \mathcal{W}^{1/r}_{(\nu+1),2}$. Repeating this process for all (ν, r) pairs it follows from Theorem 6.5 that $I_{\nu,r}(n) \downarrow 1$ with $f \in \bigcup_{\nu,r} \mathcal{W}^{1/r}_{(\nu+1),2}$.

C.8 Proof of Theorem 7.1

The iteration in (7.2) can be represented by a function $f : \mathbb{R}^N \to \mathbb{R}^N$ defined as:

$$f(\mathbf{u}) = \frac{1}{1 + 4\lambda\left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4}\right)}\mathbf{d} + \mathbf{A}\mathbf{u}.$$

Let $||\cdot||$ denote the 2-norm for vectors and the induced spectral 2-norm for matrices. Note that any row of $\mathbf{A}$ contains at most six non-zero elements with elements a_{mn} defined in (7.3). By the Perron-Frobenius theorem [190], a bound on the largest eigenvalue of $\mathbf{A}$, and hence on $||\mathbf{A}||$ can be obtained as follows:

$$\begin{aligned}
||\mathbf{A}|| &\leq \max_m \sum_n a_{mn} \\
&\leq \frac{4\lambda/\Delta x^4}{1 + 4\lambda\left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4}\right)} \\
&\quad + \frac{4\lambda/\Delta y^4}{1 + 4\lambda\left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4}\right)} \\
&\quad + \frac{4\lambda/\Delta z^4}{1 + 4\lambda\left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4}\right)} \\
&= \frac{4\lambda'}{1 + 4\lambda'}
\end{aligned}$$

$$< \quad 1$$

where $\lambda' = \lambda \left(\frac{1}{\Delta x^4} + \frac{1}{\Delta y^4} + \frac{1}{\Delta z^4} \right)$.

Therefore,

$$\begin{aligned} ||f(\mathbf{u}) - \mathbf{f}(\mathbf{v})|| &= ||\mathbf{A}(\mathbf{u} - \mathbf{v})|| \\ &\leq ||\mathbf{A}||\,||\mathbf{u} - \mathbf{v}|| \\ &< ||\mathbf{u} - \mathbf{v}||, \end{aligned}$$

where the last step follows from the fact that $||\mathbf{A}|| < 1$. Accordingly, f is a contraction map and by the Banach fixed point theorem [185] the sequence of iterates must converge to a limit point.

D

Other Applications

This appendix presents some applications of the MAPSlope-FastTrellis algorithm to data from political science and finance to demonstrate the practical utility of this procedure as a general purpose signal processing algorithm.

D.1 Presidential Approval Ratings

United States presidential job approval ratings have been published by various agencies over the years. These ratings are usually quoted as percentage values that are computed from the results of opinion polls administered to a sample of the country's population. An analysis of the approval ratings for President George W. Bush is presented by Ratkovic and Eng [73]. The result of applying the MAPSlope algorithm to a snippet from the same dataset is shown in Fig. D.1. A sudden transition in the dataset correlates with the 9/11 terrorist attack.

The raw data is assumed to be piecewise constant, which implies that the running-sum over this plot gives a piecewise linear curve. This noisy piecewise linear data can be interpreted as a "cumulative approval score" plotted as a function of time. For comparison, results from fitting a fourth order polynomial and a 3-point 1st order polynomial LOWESS filter [72] are also shown in Fig. D.1. Observe that close to the breakpoint the

Table D.1: Numerical evaluation of MAPSlope for Bush's approval ratings

	Polynomial	Lowess	FastTrellis	Smooth Optimization
Squared residual	37.61	7.8	5.13	9.84
Average # breaks	-	-	4	3.63

Residual sum squared values are calculated with respect to the raw data. Note that the average number of breakpoints is similar for both the FastTrellis and smooth optimization methods but the residual is smaller with the former. The polynomial fit is of order 4, whereas the LOWESS smoother uses a sliding window of 3 samples.

absolute residual error from the MAPSlope algorithm is the lowest. Also, MAPSlope-FastTrellis detects a change point immediately following the 9/9–11 polling period. A similar result is obtained in [73] using a non-parametric segmented spline fitting method. Additional numerical evaluation is presented in Table D.1. MAPSlope was run 1000 times with different initial guesses for the parameter values; σ^2 was drawn from $\mathcal{U}([0.05, 0.15])$ and p was drawn randomly from $\mathcal{U}([0.88, 0.93])$ A maximum limit of 6 alternating maximization iterations was used with $M = 15$ slope quantization levels. The smooth optimization algorithm was run with $\alpha = 10^{-3}$. Since there is no "ground truth" answer to the fitting problem in this application, a judgment is made using the residual squared error from the data. MAPSlope with the trellis dynamic program gives the lowest error as seen from Table D.1.

D.2 Interest Rates

Markov switching models have been used in the past to unravel piecewise constant trends in interest rate datasets. Hamilton [191] uses a 2-state Markov chain to model an economy that switches between fast and slow growth cycles. As an extension, Garcia and Perron [192] use a 3-state Markov switching model to analyze inflation adjusted quarterly interest rate data for the United States between 1961 to 1986. Assuming that

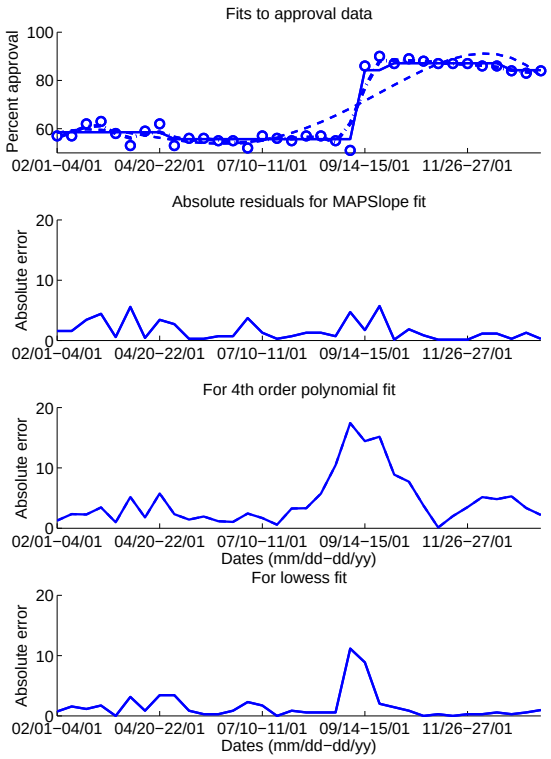


Figure D.1: Fits to Presidential approval data for President George W. Bush around 9/11. Absolute residuals obtained from MAPSlope-FastTrellis (solid line), a 4th order polynomial fit (dashed line) and a 3-point 1st order LOWESS fit (dash-dotted line) are also shown. Observe that the 4th order polynomial fit has a large error at the breakpoint. The 3-point LOWESS smoother performs slightly better than the näive polynomial fit, but still fails to capture the sharp jump.

this data follows a first-order Markov property, the MAPSlope algorithm provides an easy generalization while providing a tractable algorithm for a moderate number of quantized interest rate levels.

The result of applying the MAPSlope algorithm to the dataset presented in [74] is shown in Fig. D.2. As in [74], it is assumed that this data is piecewise constant, with occasional changes in interest rate. The raw data is integrated to obtain a cumulative interest rate plot as a function of time. This plot is assumed to be piecewise linear and

MAP estimates of the slope values are obtained. Initial parameter values are guessed from data; p is chosen close to 0.97 with the anticipation of around three regimes in the raw data vector of length 100, whereas the initial σ^2 is estimated from the sample variance after appropriate detrending. The slope values are quantized to 15 levels (which gives a resolution of about 0.067 on the unit interval).

Table D.2: Numerical evaluation of MAPSlope for the interest rate data

	Bai-Perron	FastTrellis	Smooth Optimization
MSE (Eq. (4.21))	0	0.64	1.59
Squared residual	4.32	4.74	5.37
Average # breaks	3	3	6.98

The MSE values are obtained according to (4.21) by using the Bai-Perron fit as the ground truth. Residual sum squared values are calculated with respect to the raw data. Note that the average number of breakpoints is quite high in the smooth optimization method because the quantization step often leads to a sequence of small jumps.

Fig. D.2 also shows results obtained from the Bai-Perron algorithm [193] for comparison. The output of MAPSlope-FastTrellis agrees quite well with the Bai-Perron method as seen from the plot of absolute difference. Additional numerical evaluation results are shown in Table D.2. Both FastTrellis and smooth optimization algorithms were simulated 1000 times with different initial guesses for (σ^2, p) with σ^2 drawn randomly from $\mathcal{U}([0.2, 2])$ and p drawn from $\mathcal{U}([0.90, 0.98])$. A maximum iteration limit of MAXIT $= 6$ was used. For the smooth optimization algorithm, a sequential quadratic program [128] was used, followed by quantization to the discrete set of slopes. An initial guess of 0.5 was used for all the N slopes. The smooth optimization algorithm was run with $\alpha = 10^{-3}$. It was seen that MAPSlope converged to a local stationary point $(\sigma^2, p) = (0.12, 0.97)$ in 3–4 iterations when using FastTrellis but did not converge to

a stable parameter value when the smooth optimization method was used. This is because the latter is not guaranteed to produce the optimal slope sequence, causing the hypotheses of Theorem 4.2 to be violated.

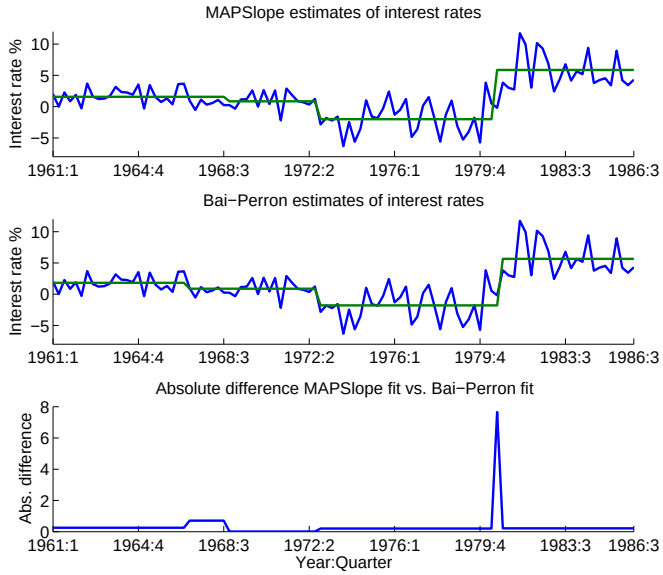


Figure D.2: Piecewise constant estimates of quarterly interest rates, data obtained from the paper by Bai and Perron [74]. The MAPSlope-FastTrellis fit was obtained using $(\sigma^2, p) = (0.12, 0.97)$ which was found to be a local convergence point of the alternating maximization scheme.