Uniform precision ultrasound strain imaging

G.M. Treece, J.E. Lindop, A.H. Gee and R.W. Prager

CUED/F-INFENG/TR 624

March 2009

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

Corresponding e-mail: gmt11@eng.cam.ac.uk

Abstract

Ultrasound strain imaging is becoming increasingly popular as a way to measure stiffness variation in soft tissue. Almost all techniques involve the estimation of a field of relative displacements between measurements of tissue undergoing different deformations. These estimates are often high resolution, but some form of smoothing is required to increase the precision, either by direct filtering or as part of the gradient estimation process. Such methods generate uniform *resolution* images, but strain quality typically varies considerably within each image, hence a trade-off is necessary between increasing precision in the low quality regions and reducing resolution in the high quality regions. We introduce a smoothing technique, developed from the nonparametric regression literature, which can avoid this trade-off by generating uniform *precision* images. In such an image, high resolution is retained in areas of high strain quality but sacrificed for the sake of increased precision in low quality areas. We contrast the algorithm with other methods on simulated, phantom and clinical data, for both 2D and 3D strain imaging. We also show how the technique can be efficiently implemented at real time rates with realistic parameters on modest hardware. Uniform precision nonparametric regression promises to be a useful tool in ultrasound strain imaging.

Contents

1 Introduction

It seems likely that some form of ultrasonic strain imaging will be adopted into routine clinical practice, within a decade, to support a still unestablished set of diagnostic tasks, primarily within the broad category of soft tissue examinations. Applications discussed in the academic literature have included detection of soft tissue tumours (Garra et al., 1997; Regner et al., 2006; Svensson and Amiras, 2006), discrimination without biopsy between complex cysts and malignant breast lesions (Barr, 2006), monitoring of atherosclerosis (de Korte et al., 1998, 2000), detection and grading of deep vein thrombosis (Emelianov et al., 2002), assessment of skin pathologies (Vogt and Ermert, 2005) and evaluation of myocardial fitness (Kaluzynski et al., 2001).

There are currently a variety of techniques for generating strain images using ultrasound, and it is not yet clear which of these techniques will be most appropriate for each of these applications. However, the majority of techniques involve the local estimation of tissue displacement by comparing radio frequency (RF) ultrasound data acquired at differing tissue deformation states. The tissue deformation can be induced in a variety of ways: in the remainder of this paper, we will focus on quasi-static ultrasound strain imaging, where the tissue is deformed by varying the contact pressure between the probe and the skin surface. However, the algorithms we develop apply equally to other strain imaging techniques. Many methods have been proposed for displacement estimation, *e.g.*, (Alam et al., 1998; Céspedes and Ophir, 1993; Céspedes et al., 1995; Lindop et al., 2007, 2008e; Lubinski et al., 1999; Maurice and Bertrand, 1999; O'Donnell et al., 1994; Pesavento et al., 1999; Pinton et al., 2006; Sumi, 1999; Viola and Walker, 2003; Zhu and Hall, 2002). Such methods typically produce high resolution displacement estimates, however the measurement quality can vary enormously across a single image, for instance due to variation in signal strength or decorrelation caused by non-axial movement.

In quasi-static strain imaging, displacement estimation is followed by gradient estimation in the axial direction. Simple differencing of consecutive samples (Ophir et al., 1991) amplifies the high-frequency components of the measurement noise. Hence differencing is often achieved by more complex techniques such as piecewise-linear least squares regression (PLLSR) (Kallel and Ophir, 1997), moving-average filtering (O'Donnell et al., 1994) and staggered strain estimation (Srinivasan et al., 2002). All such linear techniques can be interpreted as simple differencing followed by filtering with fixed kernel coefficients. Indeed, we have previously shown that, except in the case where the entire data set genuinely consists of noisy measurements from a single linear trend (in which case PLLSR is the optimal filter), simple differencing followed by filtering with a Gaussian-shaped kernel can achieve lower estimation noise than these methods at the same resolution (Lindop et al., 2008b).

Since both the displacement tracking and filtering techniques make use of kernels with fixed size, subsequent strain images have fixed resolution but variable quality. However, this variation can be quantified, since it is straightforward to obtain a reasonable estimate of the precision (inverse of measurement variance) of each measurement (Lindop et al., 2008a). Strain images require some form of normalisation to convert the strain into a displayable range, and to reduce variation that is simply a result of variation in the applied stress (Lindop et al., 2008c). The precision of the displayed strain value depends both on the displacement estimation precision and on the normalisation value used at each point in the image. Both of these factors can vary within each image, leading to large variations in precision which can make strain images hard to interpret.

In order to prevent confusion due to the display of low precision strain data, images are often suppressed once the overall precision falls below a fixed threshold (Jiang et al., 2007). However, strain images with low overall precision can still contain high precision regions, and this is exploited by techniques which combine multiple images, using local strain precision information to ensure the

best data in each image contributes more to the final result. Such data still contains regions of low precision, but these can be masked by use of a suitable colour wash (Lindop et al., 2008c).

We present here a method for producing strain images with uniform *precision* and varying *resolution*, rather than uniform *resolution* and varying *precision*. Such images may be easier to interpret: lack of precision in strain images leads to regions which falsely appear to have strong fine-scale stiffness variation, whereas lack of resolution leads to high levels of blurring, which is more easily interpreted. In this case, a colour wash can be additionally used to suppress areas with very low resolution (rather than low precision as before). Whether this approach is indeed better is clearly somewhat subjective, hence the results are mostly presented in visual form, so readers can judge for themselves.

In Section 2 we describe the principle behind non-uniform smoothing of strain data, followed by details of our implementation, since computational issues are important in the context of real-time applications. Section 3 contains an analysis of the resolution and precision of the subsequent strain estimates, leading to a formulation for uniform output precision. In Section 4, we compare the technique with PLLSR and Gaussian filtering, including simulations, phantom studies and clinical examples, for both 2D and 3D strain data. General conclusions are drawn from these results in Section 5.

2 Method

To provide a more general framework for smoothing strain images, we follow the roughness penalty approach to nonparametric regression (NPR) (Green and Silverman, 2004). It should be noted, however, that the resulting equations can be arrived at from a variety of directions, for instance variable-kernel smoothing (Silverman, 1984), which lead to different interpretations of largely the same parameters. Since we are considering regression in the context of image filtering, we want an approach which allows user control over the *extent* of filtering (equivalent to the window length in PLLSR, for example) whilst automating the local smoothing properties. NPR is a good candidate for this, since it depends on settings which may be thought of as data weights (which can be automatically chosen for uniform precision) and a smoothing strength (which can be controlled to adjust the level of precision).

In principle, since NPR is a linear operation, we could apply it either before differencing the displacement data, or after, on the strain data. However, there are two key reasons in practice why it makes more sense to apply it to strain data:

- Strain data varies with the amount of applied stress. Since we really want to visualise stiffness, strain data needs some form of normalisation before it can be usefully displayed, loosely equivalent to dividing through by an estimate of applied stress. Hence this also results in a change in data precision: low stress areas are then correctly identified as having lower precision even though the displacement precision may have been high[1]. Applying NPR at this stage allows us to identify such regions correctly, resulting in far better images if the applied stress was highly non-uniform due to poor probe movement.

- In order to produce a high quality display, strain data is often persisted over a sequence of images. This persistence must be over normalised strain in order to ensure the image levels

[1]This is easier to see in the limit of no applied stress — in this case the displacement precision is very high, since there is no deformation, but the displayed precision *must* be very low, since it is not possible to measure stiffness if there was no deformation at all.

are consistent. If we filter the data *before* persistence, we would be persisting images with inconsistent resolution, and this produces very undesirable results. For correct persistence, we need to keep the resolution constant, and only perform variable resolution NPR on the final persisted data, just before display.

That said, for very good quality data, with no persistence and an ideal probe motion which does not require normalisation, slightly better results are possible by applying NPR to displacement data. This topic is discussed in more detail by Lindop et al. (2008d).

2.1 Principle

NPR is described most simply as an optimisation problem. A regression curve (in one dimension) or a surface (in two or more dimensions) is found by minimising a particular cost function (Whittaker, 1923). A first group of terms, known as the data penalty, expresses the cost of a poor fit between raw and regressed strain. A second group of terms, known as the roughness penalty, ensures that a relatively low cost is incurred by simple surfaces, such as planes, whereas a relatively high cost is incurred by complicated surfaces, such as those containing large numbers of sharp peaks and troughs. The continuous form of the roughness penalty is an integral over the entire surface of the square of appropriate surface derivatives. We consider two particular cost functions:

$$C_1 = \iint w_1 \left(\hat{\mathrm{s}} - \acute{\mathrm{s}}_1\right)^2 + r_1 \left[\left(\frac{\partial \acute{\mathrm{s}}_1}{\partial x} \right)^2 + \left(\frac{\partial \acute{\mathrm{s}}_1}{\partial y} \right)^2 \right] \mathrm{d}x \, \mathrm{d}y \tag{1}$$

$$C_2 = \iint w_2 \left(\hat{\mathrm{s}} - \acute{\mathrm{s}}_2\right)^2 + r_2 \left[\left(\frac{\partial^2 \acute{\mathrm{s}}_2}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \acute{\mathrm{s}}_2}{\partial y^2} \right)^2 \right] \mathrm{d}x \, \mathrm{d}y \tag{2}$$

where x and y denote lateral and axial distance (it is assumed that strain is measured in the axial direction), $\hat{\mathrm{s}}$ and $\acute{\mathrm{s}}$ are the raw and regressed strain data respectively, which are multiplied by data weights, w. For C_1, The roughness penalty consists of squared partial first derivatives in the x and y directions, whereas C_2 contains squared partial second derivatives. Hence C_1 penalises strain data which is not reasonably constant, whilst C_2 penalises strain data which is not reasonably smooth. In both cases r, the smoothing strength, controls the degree of 'reasonableness' in the previous statements[2].

In practice, we do not have a continuous description of the raw strain data, and we do not require a continuous function for the regressed strain. Our aim is simply to replace the raw data with a filtered version of the data. In this case, a discretised form of eqs. (1) and (2) can be expressed in matrix form:

$$C \approx (\acute{\mathbf{s}} - \hat{\mathbf{s}})^T \mathbf{W} (\acute{\mathbf{s}} - \hat{\mathbf{s}}) + r \acute{\mathbf{s}}^T \mathbf{M}^T \mathbf{M} \acute{\mathbf{s}} \tag{3}$$

where $\mathbf{W}$ is now a diagonal matrix of data weighting factors, with one entry per data point. $\hat{\mathbf{s}}$ and $\acute{\mathbf{s}}$ are vectors listing the raw and regressed strain data respectively. For $C \equiv C_1$, $\mathbf{M} \equiv \mathbf{M_1}$ is a matrix which extracts every first difference, in both x and y directions, and this is weighted by the constant $r \equiv r_1$. For $C \equiv C_2$, $\mathbf{M} \equiv \mathbf{M_2}$ lists every second difference; further details are given in Appendix A.

A formula for the regression surface is found by setting $\nabla C = \mathbf{0}$, differentiating with respect to the value at every point on the regression surface, $\acute{\mathbf{s}}$, which gives

$$(\mathbf{W} + r\mathbf{M}^T\mathbf{M})\acute{\mathbf{s}} = \mathbf{W}\hat{\mathbf{s}}. \tag{4}$$

[2] If the image scale is not the same in each direction, different values of r are used to weight each partial derivative, in order to maintain equal smoothing.

The form of the regressed data $\acute{\mathbf{s}}$ is therefore dependent on the data penalty $\mathbf{W}$ and roughness penalty $\mathbf{M}$. Setting $\mathbf{W}$ to a constant multiplied by the identity matrix results in a simple sum-of-squared error data penalty and will produce uniform smoothing, but allowing the components of the leading diagonal of $\mathbf{W}$ to vary will result in non-uniform smoothing. The roughness penalty $\mathbf{M}$ should reflect how we expect the strain data to behave: $\mathbf{M} \equiv \mathbf{M_1}$ represents strain which is approximately constant, at least locally, and $\mathbf{M} \equiv \mathbf{M_2}$ represents strain which is expected to vary smoothly. Both of these constraints might be useful for real data, and both are tested in Section 4.

2.2 Implementation

Solving eq. (4) involves inversion of the $(\mathbf{W} + r\mathbf{M}^T\mathbf{M})$ matrix combination either by direct or iterative means. For non-uniform smoothing, the data weights $\mathbf{W}$ themselves depend on the data, and hence this inversion must be repeated every time an image is filtered. Efficient NPR in 1D was first described by Reinsch (1967). In this case $(\mathbf{W} + r_1{\mathbf{M_1}}^T\mathbf{M_1})$ is a symmetric positive definite tri-diagonal matrix and $(\mathbf{W} + r_2{\mathbf{M_2}}^T\mathbf{M_2})$ is penta-diagonal, and in both cases direct inversion is easily achieved by either the Thomas algorithm or band-limited Cholesky decomposition (George and Liu, 1981).

In 2D, however, the matrices are of size $N_xN_y \times N_xN_y$, where N_x, N_y is the size of each image, and the mid-band for $\mathbf{M_1}$ is of width $2N_x+1$ (or $2N_y+1$, dependent on the ordering of the data in $\acute{\mathbf{s}}$) and for $\mathbf{M_2}$ of width $4N_x + 1$. This makes direct inversion for 2D or 3D data impractical for a real-time system, and we need instead to use an iterative method. In Section 4 we compare the performance of the point Gauss-Seidel method (Varga, 2000), the method of conjugate gradients (Press et al., 2002) and both the multigrid and full multigrid techniques, described in more detail in Appendix B.

Most iterative techniques benefit from a good initial guess at the solution. A pseudo-1D method can be used to provide such a guess at very low computational cost, by performing smoothing in each of the directions independently:

$$\begin{aligned} (\mathbf{W} + r\mathbf{M_x})\acute{\mathbf{s}}_x &= \mathbf{W}\hat{\mathbf{s}} \\ (\mathbf{W} + r\mathbf{M_y})\acute{\mathbf{s}}_{1D} &= \mathbf{W}\acute{\mathbf{s}}_x \end{aligned} \tag{5}$$

where $\acute{\mathbf{s}}_{1D}$ is the pseudo-1D solution and $\mathbf{M_x}$ and $\mathbf{M_y}$ represent roughness penalties in each direction independently, such that $\mathbf{M_x} + \mathbf{M_y} = \mathbf{M}^T\mathbf{M}$. Expanding eq. (5) and substituting for $\mathbf{M}$ gives

$$\begin{aligned} (\mathbf{W} + r\mathbf{M_x})\mathbf{W}^{-1}(\mathbf{W} + r\mathbf{M_y})\acute{\mathbf{s}}_{1D} &= \mathbf{W}\acute{\mathbf{s}} \\ \therefore \quad (\mathbf{W} + r\mathbf{M}^T\mathbf{M} + r^2\mathbf{M_x}\mathbf{W}^{-1}\mathbf{M_y})\acute{\mathbf{s}}_{1D} &= \mathbf{W}\acute{\mathbf{s}} \end{aligned} \tag{6}$$

By comparison to eq. (4), eq. (6) contains the additional term $r^2\mathbf{M_x}\mathbf{W}^{-1}\mathbf{M_y}$. This tends to generate small, high frequency errors in the regressed strain data, leaving the low frequency content remarkably close to the full 2D solution. Since iterative schemes are typically much better at reducing high frequency than low frequency errors, using $\acute{\mathbf{s}}_{1D}$ as an initial guess for $\acute{\mathbf{s}}$ is particularly successful.

Equation (4) and the pseudo-1D initialisation in eq (5) have been presented in 2D, but they are both trivially extensible to 3D strain data.

3 Analysis

We have seen that NPR can be used to provide non-uniform smoothing of the strain data, but we would like to know how to set the data weight $\mathbf{W}$ so as to provide exactly the right non-uniform

smoothing to generate a uniform precision image. To do this, we examine how the resolution varies with varying data weight w, assuming that the underlying strain is largely constant, and the raw strain estimates are independent and of high resolution. We then briefly consider the effects of violating these assumptions in practice.

3.1 Resolution

Since NPR is a linear operation, the regressed strain data at each point can be calculated as a weighted sum of some kernel coefficients with the original data. These kernel coefficients are different for every data point in the regressed data set, however if we assume uniform smoothing (i.e. $\mathbf{W} = w\mathbf{I}$), with w not too large, and only consider data away from the edges of the image, we can model NPR as a convolution with an effective kernel $H(\omega_x, \omega_y)$ (Silverman, 1984). Now let $\hat{\mathrm{s}}$ be a continuous raw strain measurement with spatial frequencies ω_x, ω_y:

$$\hat{\mathrm{s}} = \cos(\omega_x x + \omega_y y)$$

hence the regressed strain data is approximated by

$$\acute{\mathrm{s}} = |H| \cos(\omega_x x + \omega_y y + \angle H)$$

where we abbreviate $H \equiv H(\omega_x, \omega_y)$ for concision. Substituting these terms into eq. (2) gives

$$C_2 = \iint \left(w(1 - |H|)^2 + r_2(\omega_x^4 + \omega_y^4) \, |H|^2 \right) \cos^2(\omega_x x + \omega_y y + \angle H) \, \mathrm{d}x \, \mathrm{d}y,$$

and differentiating with respect to $|H|$ leads to the kernel values at minimum C_2:

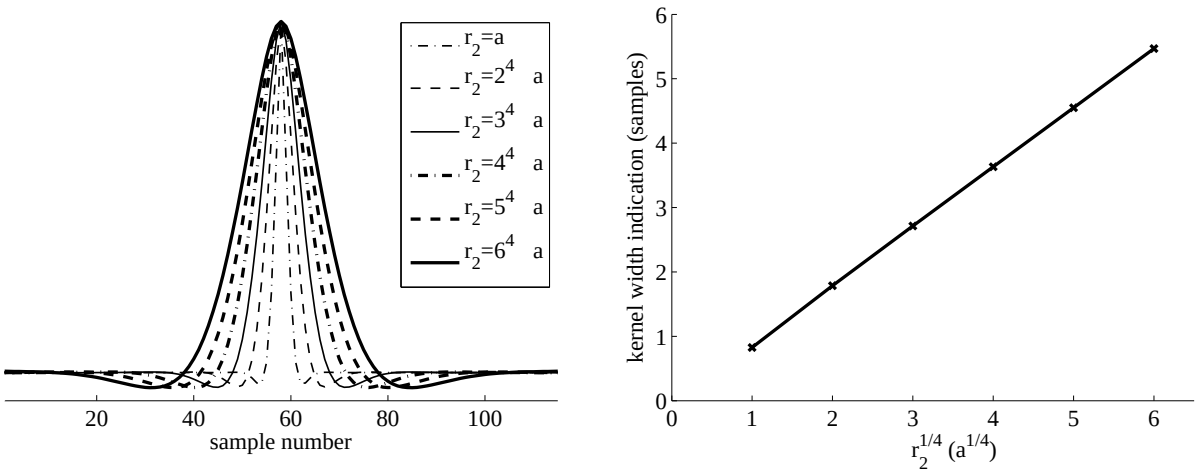
$$\begin{aligned} \frac{\partial C_2}{\partial |H|} = 0 &= w(|H| - 1) + r_2(\omega_x^4 + \omega_y^4) \, |H| \\ \therefore |H(\omega_x, \omega_y)| &= \frac{1}{1 + \frac{r_2}{w} \left(\omega_x^4 + \omega_y^4 \right)} \end{aligned} \tag{7}$$

The equivalent result for C_1 is:

$$|H(\omega_x, \omega_y)| = \frac{1}{1 + \frac{r_1}{w} \left(\omega_x^2 + \omega_y^2 \right)} \tag{8}$$

The equivalent kernel for NPR is therefore a low pass filter, with cut-off frequency in both the axial and lateral directions of $(w/r_1)^{(1/2)}$ for NPR with C_1 and $(w/r_2)^{(1/4)}$ for NPR with C_2. The resolution in each direction will be proportional to $(r_1/w)^{(1/2)}$ for C_1 and $(r_2/w)^{(1/4)}$ for C_2, assuming that NPR is responsible for the resolving limit, i.e. the raw strain data before NPR has a sufficiently high resolution.

This relationship is empirically verified in Fig. 1, with plots showing the equivalent kernel over the axial dimension for six examples of different smoothing strengths, all using C_2. The smoothing strengths are expressed in terms of a constant scale factor, a. The coefficients are found by calculating the response to an impulse in the centre of the image, with a uniform data weight. An indication of kernel width can be derived from these kernel coefficients, and this is plotted against $r_2^{1/4}$ to confirm the expected relationship.



(a) (b)

Figure 1: C_2 NPR resolution with uniform smoothing. (a) Normalised effective NPR kernels over the axial dimension for various smoothing strengths, r_2. (b) Kernel width against $r_2^{1/4}$, confirming the expected relationship.

3.2 Precision

Assuming the raw strain data contains independent measurements from the same relatively homogeneous strain distribution, the regressed strain data precision will scale with the size of the equivalent NPR kernel, which is effectively acting to average the raw strain values[3]. The 2D kernel has a size proportional to the square of the 1D kernel width, shown in Fig. 1. For raw data with precision p, the precision of the regressed data for NPR using C_2 will be $(r_2/w)^{(1/2)}p$, whereas for C_1 it will be $(r_1/w)p$.

This implies, for 2D kernels, that we can keep the precision of the regressed data constant by setting each of the diagonal elements of $\mathbf{W}$ to a value $w(x, y)$ related to the precision $p(x, y)$ of the corresponding raw data. For NPR with C_1, $w(x, y) = p(x, y)$ and with C_2, $w(x, y) = p(x, y)^2$, in which case the precision of the entire regressed image will scale with r_1 and ${r_2}^{(1/2)}$ respectively.

A similar analysis in 3D shows that we should set $w(x, y) = p(x, y)^{(2/3)}$ for C_1 NPR and $w(x, y) = p(x, y)^{(4/3)}$ for C_2 NPR to maintain uniform precision in this case.

3.3 Deviations from the simple model

Real strain data is of course not homogeneous, and neighbouring strain measurements are unlikely to be completely independent of each other. We briefly consider here some of the effects of these violations of the assumptions we have been making in the previous sections.

- We assume that the windows used for estimating each displacement are small, which has the effect of ensuring that the raw resolution is as high as possible. Since we are attempting to apply a filter which is optimal in the sense of minimising a defined cost, we want to avoid any other filtering which will presumably be less optimal. In addition, clearly NPR can only improve precision in the data by reducing the resolution, hence it makes sense to have as high a

[3] Applying NPR to smooth displacement data before taking differences produces a slightly different result, which is analysed in Lindop et al. (2008d).

resolution as possible in the raw data. In practice, the size of displacement tracking windows is set just large enough to avoid major 'peak-hopping' errors when combined with a good tracking strategy (Chen et al., 2009).

- Displacement window overlap will introduce dependencies between displacement estimates, which will affect the uniformity of precision of the regressed data. In practice a low degree of overlap seems to have little visual effect on the regressed data.

- Since strain is calculated by differencing the displacement data, there is inevitably some dependence between neighbouring strain estimates. It would be possible to include this in our analysis by setting off-diagonal elements of $\mathbf{W}$ to the appropriate covariance between the strain estimates, but we do not investigate this possibility here.

- We assume that a local estimate of the raw strain precision p is available: clearly NPR can only generate a uniform precision image where the raw precision is known. In our experience, raw precision estimates based on correlation and on residual phase variance (see Appendix C) both generate useful NPR images.

4 Experiments

We test NPR with both C_1 and C_2 cost functions, in 2D and 3D, on simulated, phantom and clinical data. PLLSR and precision-weighted Gaussian filtering are used as comparisons, though these are clearly uniform resolution rather than uniform precision filters. In every test, displacement measurements are calculated using Weighted Phase Separation (WPS) (Lindop et al., 2008e), with the tracking strategy described by Treece et al. (2006). Displacement locations are adjusted using Amplitude Modulation Correction (Lindop et al., 2007) and interpolated to a regular grid as described by Treece et al. (2008)[4]. Displacement precision is calculated from the reciprocal of a weighted variance of phase differences between pre- and post-deformation RF-windows, see Appendix C.

2D equal aspect ratio windows were used for displacement tracking, of length between 6 and 8 cycles of the ultrasound probe centre frequency. Window overlap was typically between 10 and 20 percent of the window length.

For NPR and precision-weighted Gaussian filtering, strain is calculated by differencing adjacent displacement estimates, and the precision adjusted accordingly. The strain data is normalised to account for variations in probe movement, and the precision updated to affect this normalisation. For the clinical examples only, the normalised strain data is persisted with a per-pixel weighting across a few frames to improve the raw precision. These normalisation and persistence operations are as described by Lindop et al. (2008c). Filtering is then applied to the persisted, normalised strain data, immediately prior to display.

PLLSR is itself a technique for calculating strain from displacement data, which in doing so incorporates some filtering of the data. Hence this is applied to the displacement data, and the precision values updated appropriately. Normalisation, persistence (for the clinical examples) and display of the strain data are then performed without any further filtering.

Normalised strain images are displayed in grey, with black representing zero strain, and white representing twice the average image strain. A red colour wash is used to represent regions of low precision (for the raw, PLLSR and Gaussian filtered images) or low resolution (for the NPR images).

[4]This step only results in a small improvement in displacement accuracy, since we are using fairly small windows for displacement tracking.

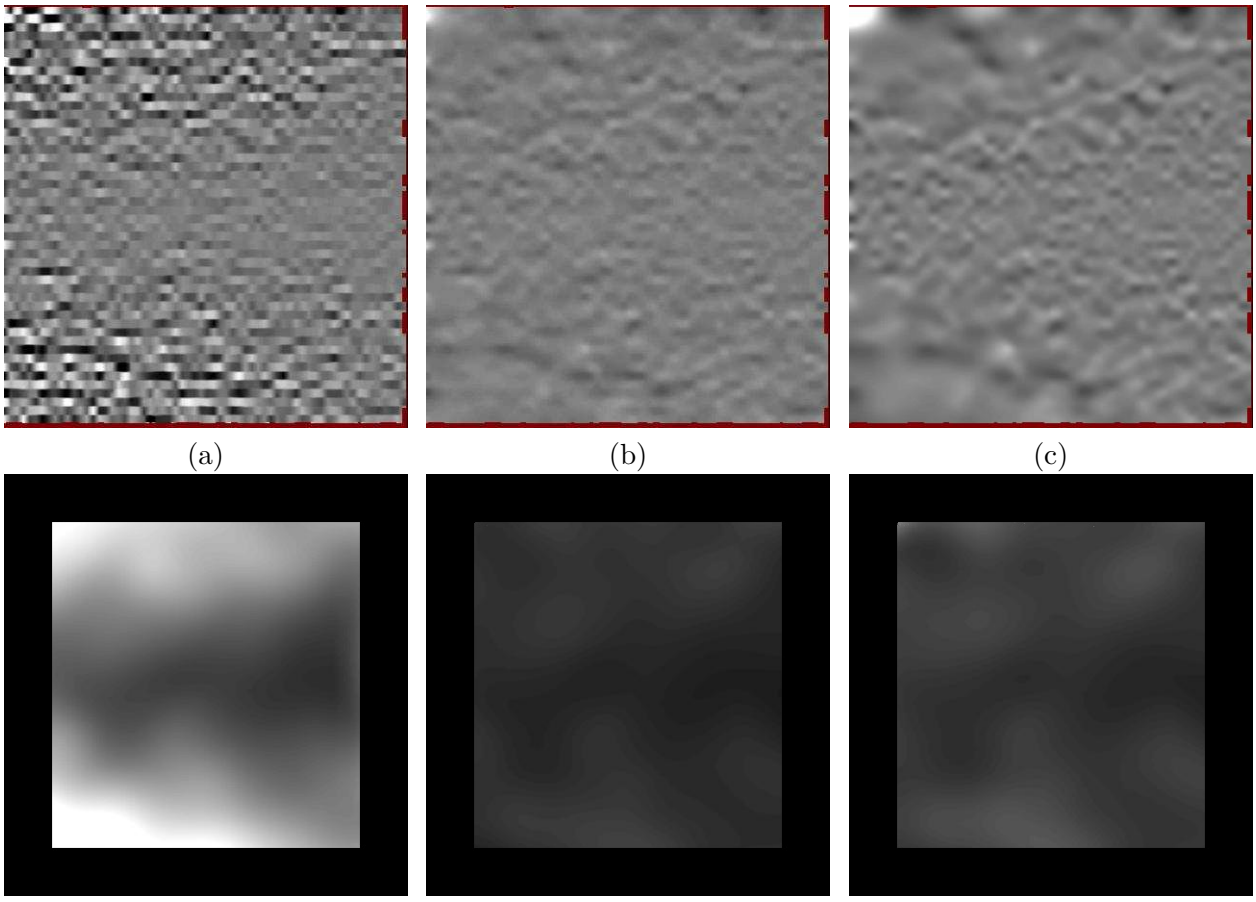


Figure 2: NPR precision with non-uniform smoothing. A simulated strain sequence is constructed with SNR which is varied laterally by adjusting scatterer density, and axially due to the single transmit and receive focus. (a) Raw high resolution strain image, with (d) corresponding precision image. (b) C_1 NPR strain image with (e) corresponding precision image. (c) C_2 NPR strain image with (f) corresponding precision image. The precision images are the result of Gaussian-filtering the squared errors.

The threshold for this colour wash varies for each of the figures in order to reveal pertinent features of the underlying filtered strain distribution.

4.1 Simulation studies

Simulated data were produced using Field II (Jensen, 1996) with parameters to mimic a 5-10MHz linear array probe with 127 elements. Pre-deformation scatterer fields representing tissue with uniform echogenicity were constructed with 10^6 scatterers distributed throughout a $50 \times 50 \times 6$ mm volume, scanned over a 40×40 mm image region.

Figure 2 contains results for a simulation of a homogeneous material with 1% strain. The signal-to-noise ratio varies throughout the data due to a lateral increase in scatterer density from left to right, and a single axial transmit and receive focus at the vertical centre of the image[5]. Local strain

[5]This data set was previously used to test a parametric method for non-uniform smoothing: see Fig. 11 in (Lindop

precision is measured by filtering the squared strain error with a Gaussian kernel, in order to visualise the extent of variation. Precision images are grey, with black representing high precision and white low precision, on a consistent scale over all of the figures.

The non-uniform precision of the raw strain data is visible in Fig. 2(a) and confirmed in the precision image of Fig. 2(d). Uniform precision NPR is then applied, using both C_1 and C_2, in order to improve the precision of the whole image to match the high precision region at the centre right of the raw strain image. Although some variation in precision is still evident in Figs. 2(e) and (f), it is clearly far more uniform here than in the original raw strain data. The filtered strain images in Fig. 2(b) and (c) show how this has been achieved in each case. As expected, NPR using C_1 imposes homogeneity in the data, but not smoothness, which results in a uniform strain image in which fine scale features are still visible. NPR using C_2 imposes smoothness but not homogeneity, so the strain image is indeed smoother, if only slightly, but there is also a greater range of strain values than with C_1.

Figure 3 contains results for a simulation with diagonal bands of width 2 mm and zero axial strain separated by a band of width 2 mm at the 2% background strain. White noise was added to reduce the SNR to 20 dB.

The raw strain image in Fig. 3(a) contains good strain measurements at the vertical centre where the ultrasound signal is well focused, but poor measurements at the top and bottom of the image. Using either PLLSR (Fig. 3(b)) or Gaussian filtering (Fig. 3(c)) improves the precision, but resolution is lost at the centre of the image before the precision at the top and bottom is improved sufficiently to clearly see the detail. In contrast, NPR (Fig. 3(d) to (f)) is able to smooth the low precision image areas without unduly affecting the resolution at the centre of the image. As with Fig. 2(c) and (e), the differences in style between C_1 and C_2 NPR are clearly visible in the strain images.

Figure 3(f) is an example of NPR with C_2 applied to the displacement data rather than the strain data. This is a situation where we can work directly with the displacement data, since the deformation is ideal, and the data is of sufficient quality not to need any persistence over multiple strain images. With simulated data such as this, NPR performs slightly better on displacements than on strain.

4.2 Phantom studies

Several test objects were scanned using a Dynamic Imaging Diasus ultrasound machine[6] with a 5-10 MHz probe focused at 20 mm depth. Frames consisted of 127 A-lines at 0.3 mm pitch, sampled at 66.7 MHz using a Gage[7] CompuScope 14200 analogue-to-digital converter. The spacing of RF-samples was 0.0115 mm, assuming a sound speed of $1540\,\mathrm{ms}^{-1}$.

Two phantoms were used. The first was a simple in-house phantom constructed from half an olive embedded in agar mixed with aluminium oxide powder. This has some advantages in that the object of interest is not spherical, and the background scattering is less regular than commercial phantoms, though the material properties are unknown. The second phantom was a commercially available breast biopsy phantom, CIRS model 052A[8], containing randomly positioned stiff inclusions which were also visible in ultrasound B-mode images. Raw strain data for scans of the half-olive phantom contained 127×291 samples, for the CIRS phantom the data contained 127×437 samples.

et al., 2008a).

[6] Dynamic Imaging Ltd., no longer in business.

[7] Gage Applied Technologies, Illinois, USA http://www.gage-applied.com.

[8] Computerized Imaging Reference Systems, Inc, Virginia, USA http://www.cirsinc.com.

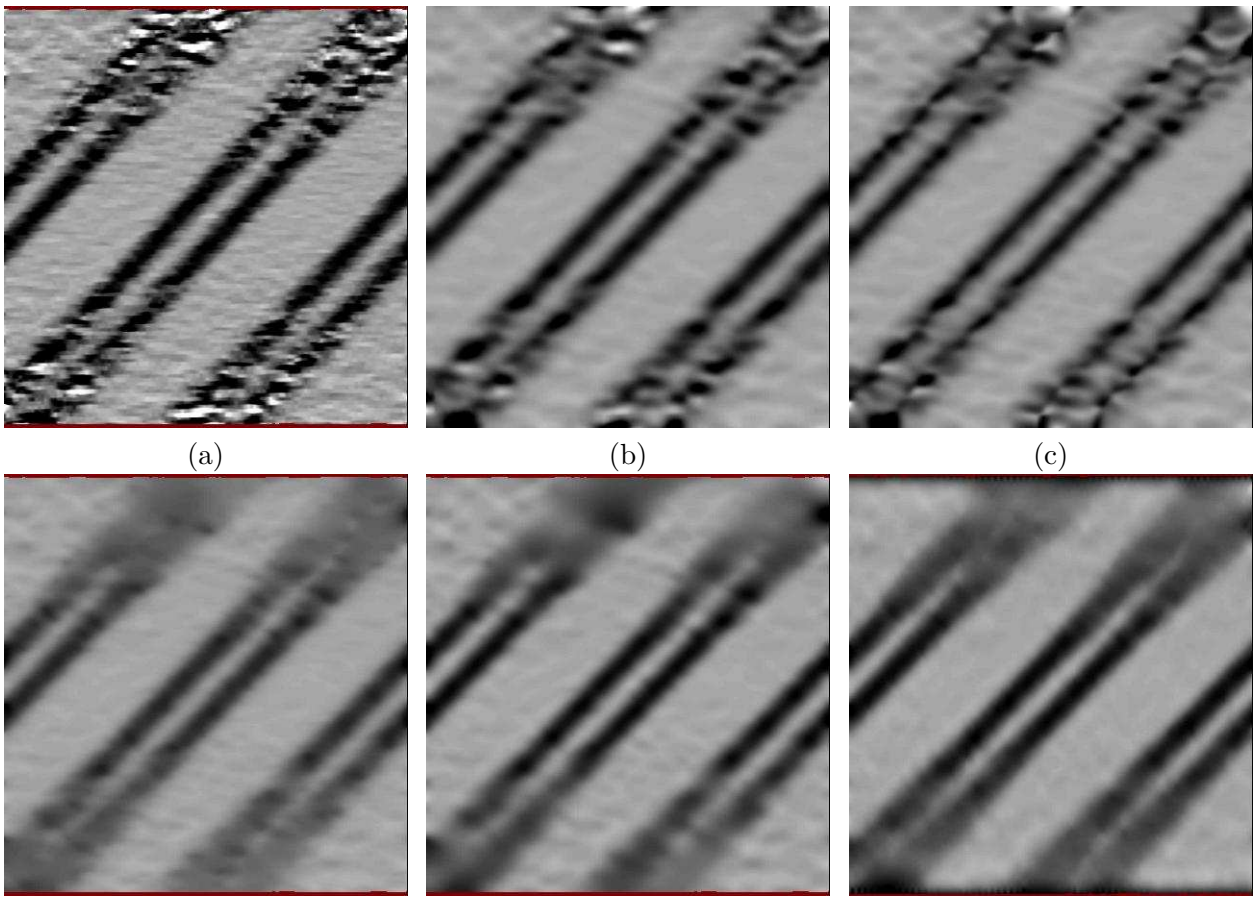


Figure 3: Simulated strain data with diagonal features. (a) Raw high resolution strain data. The remaining images are filtered using (b) PLLSR, (c) Gaussian filtering, (d) C_1 NPR, (e) C_2 NPR and (f) C_2 NPR applied to displacement, rather than strain, data. In each case the smoothing strengths have been manually optimised to give the best balance between good resolution at the centre and reasonable precision at the top and bottom.

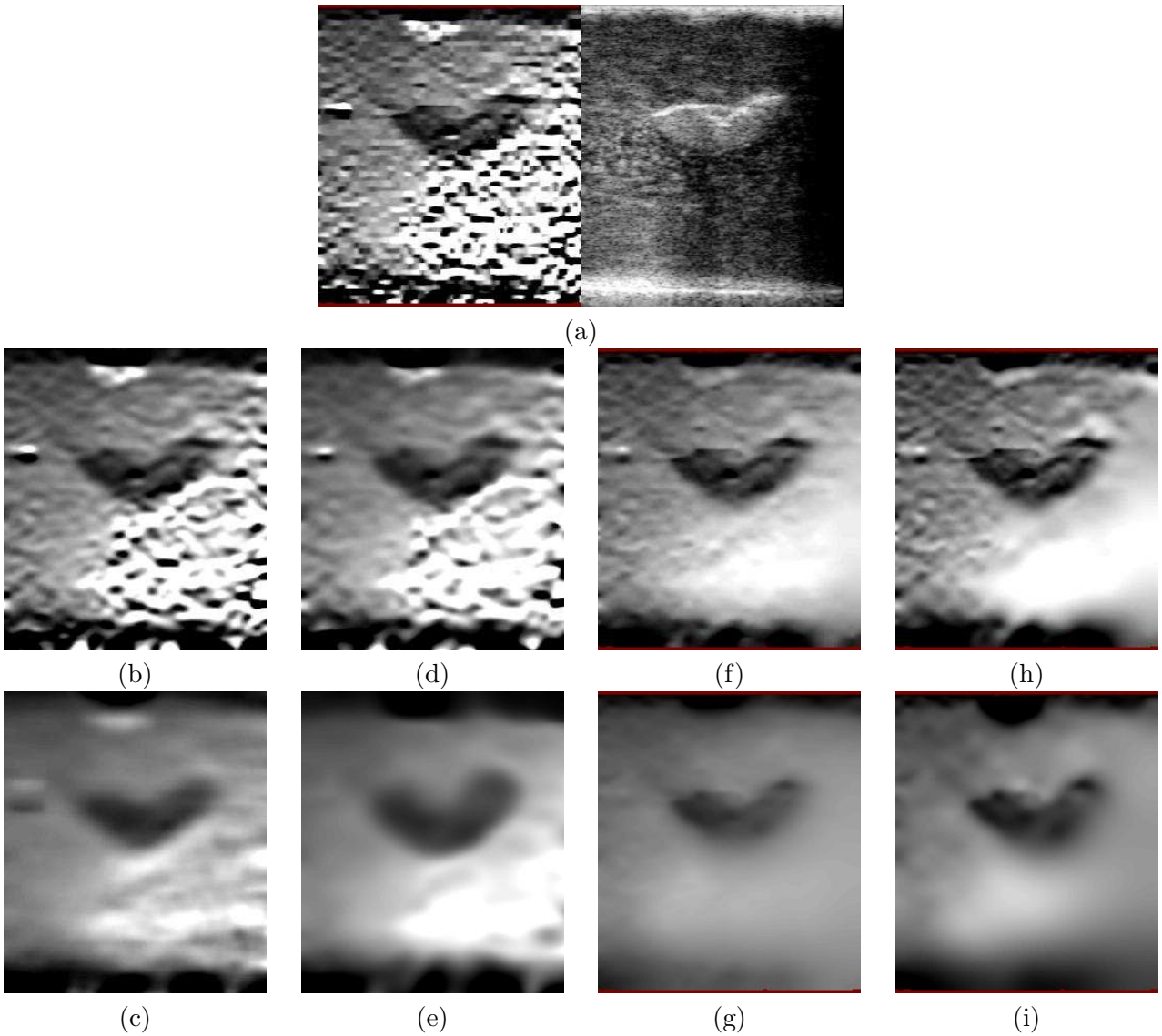


Figure 4: Strain images of the half-olive phantom. (a) The raw high resolution strain image and associated B-mode image, showing poor signal quality regions at the left and bottom. The remaining images show lightly filtered (middle row) and heavily filtered (bottom row) versions of the strain data. (b) and (c) use PLLSR, (d) and (e) use Gaussian filtering, (f) and (g) use C_1 NPR, (h) and (i) use C_2 NPR.

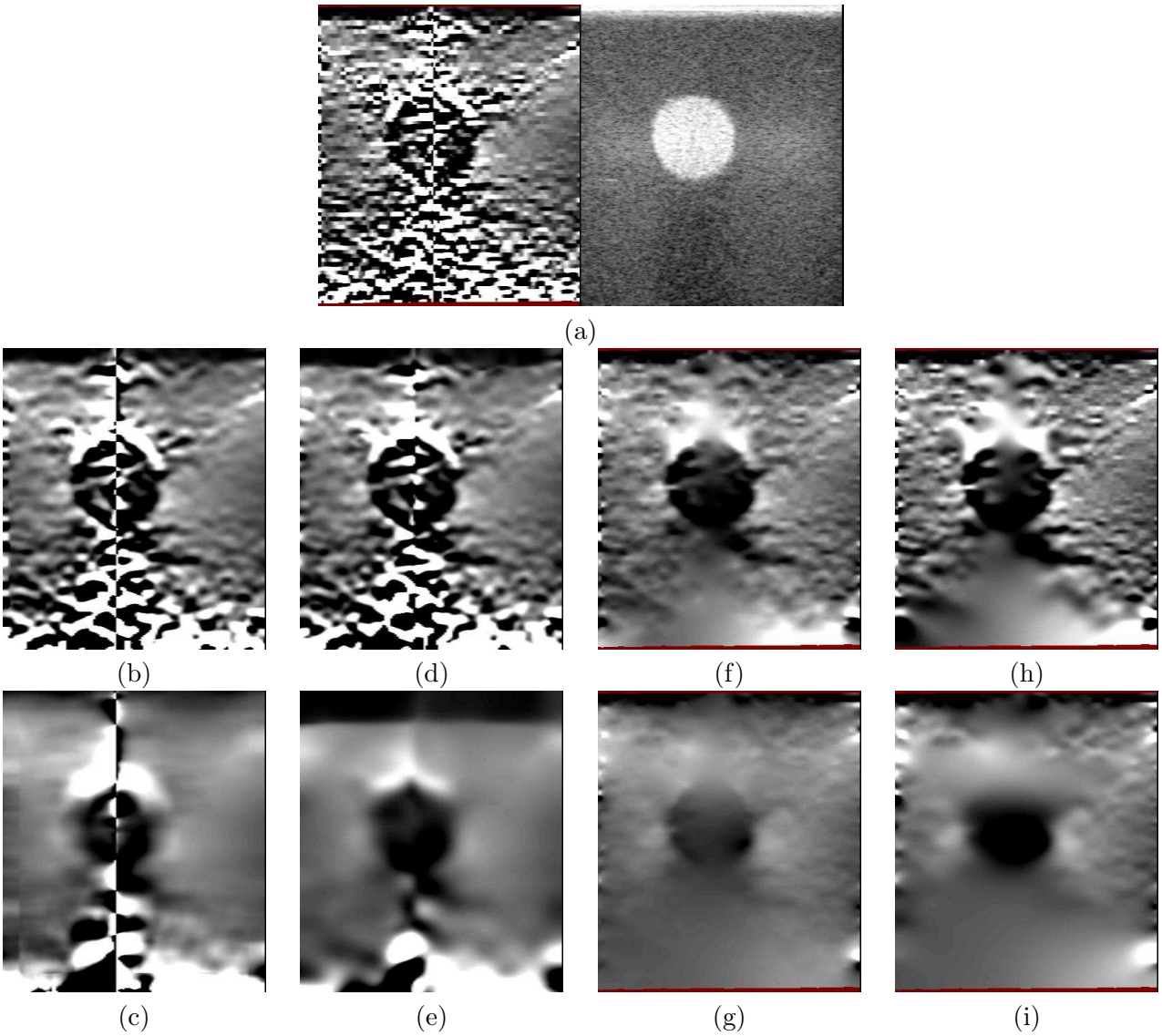


Figure 5: Strain images of the CIRS breast biopsy phantom. (a) The raw high resolution strain image and associated B-mode image, showing good B-mode signal quality, but poor strain quality in the centre due to inappropriate rotational motion of the probe. The remaining images show lightly filtered (middle row) and heavily filtered (bottom row) versions of the strain data. (b) and (c) use PLLSR, (d) and (e) use Gaussian filtering, (f) and (g) use C_1 NPR, (h) and (i) use C_2 NPR.

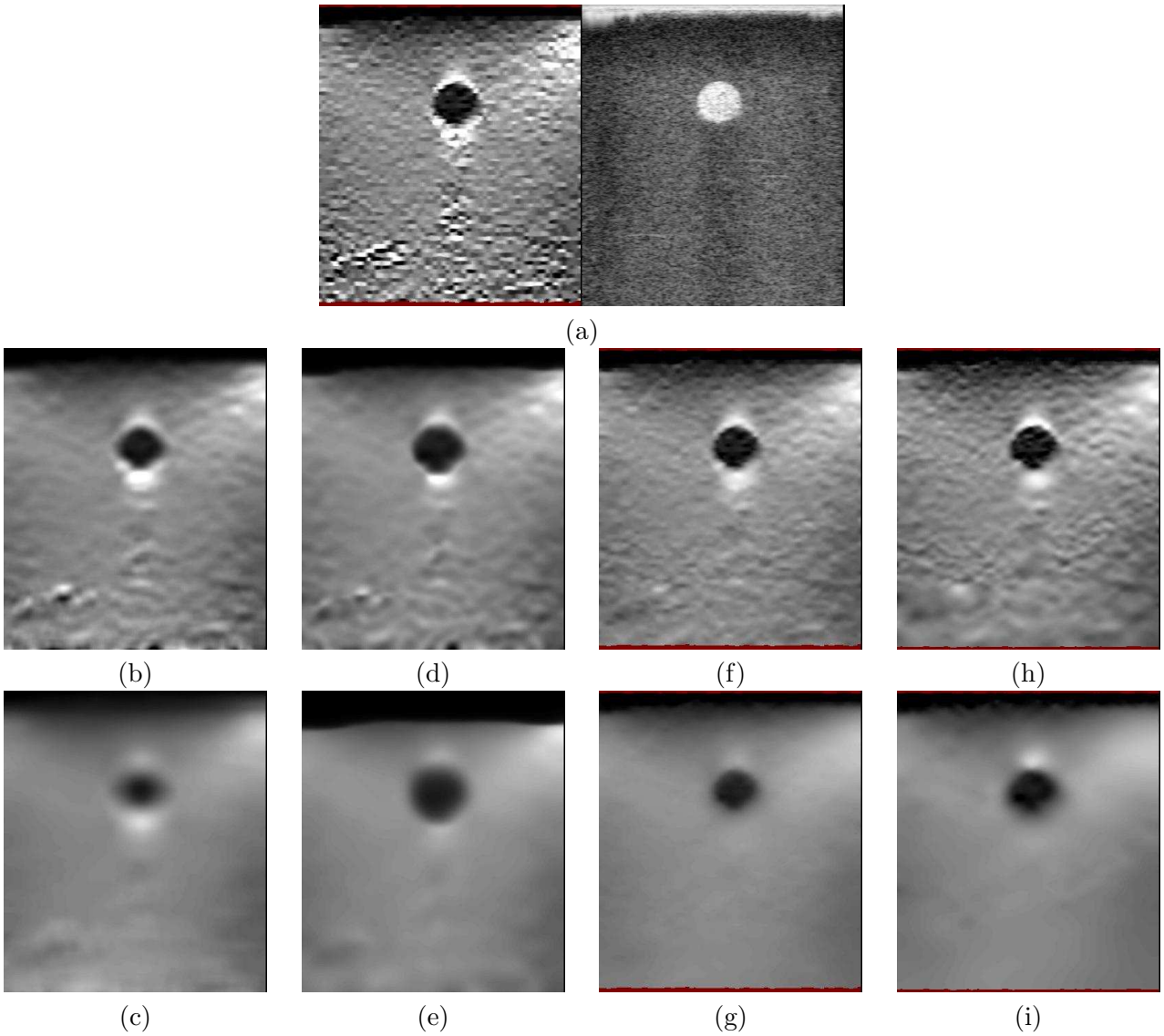


Figure 6: Strain images of the CIRS breast biopsy phantom. (a) The raw high resolution strain image and associated B-mode image, showing good B-mode and strain signal quality. The remaining images show lightly filtered (middle row) and heavily filtered (bottom row) versions of the strain data. (b) and (c) use PLLSR, (d) and (e) use Gaussian filtering, (f) and (g) use C_1 NPR, (h) and (i) use C_2 NPR.

Three scanning sequences are shown, two of the CIRS phantom, and one of the half-olive phantom, corresponding to three substantially different noise scenarios in freehand quasi-static strain imaging. In Fig 4, the half olive is scanned with a predominantly axial movement of the ultrasound probe, however, the probe contact with the phantom is poor at the right of the images, resulting in very low ultrasound signal strength in this region. The olive is also highly attenuating, and this generates severe shadowing artifacts in the inferior region. In Fig 5, the ultrasound signal strength is good throughout the data, however, the probe movement is predominantly rotational, about an axis orthogonal to the image plane roughly at the top centre of the images. This results in a high degree of variation in the applied stress field, and no axial strain in a vertical line down the centre of the image. In contrast, Fig. 6 is a scan with good ultrasound signal strength and good probe motion.

Each of these scans is processed with the four filters (PLLSR, precision-weighted Gaussian, NPR with C_1 and NPR with C_2), and two levels of filtering. Lighter filtering was set at a level, for each filter, which preserved as much of the raw image resolution as possible whilst presenting a useful image. Heavier filtering was set at a level deemed to be the highest which could sensibly be applied without over-distorting the object of interest. These levels are clearly subjective: the images are intended to give insight into the filtering properties at extreme filtering settings.

Figure 4(b) and (c) show that PLLSR can generate unexpected artifacts in the presence of noise. Some noise features in the raw data (for instance the bright-dark band to the left of the olive) become spread over a much wider area after filtering. Similar effects in the noisy region at the bottom-right of the heavily filtered image contribute to a presentation of the noise which is in some places smooth and in others has quite strong features: these could easily be misinterpreted as genuine strain discontinuities. Gaussian filtering in Fig. 4(d) and (e) performs better in this region, but at the cost of significant blurring of the olive, making it appear larger than in reality. Neither PLLSR nor Gaussian filtering suppress the noise sufficiently with light filtering. The NPR results of Fig. 4(f) to (i) all manage to preserve the important detail in the raw strain image whilst suppressing the noisy regions very successfully. Light filtering preserves the raw resolution over the olive but smooths the noise, and even with heavy filtering, the shape and size of the olive are largely preserved. Heavy filtering with C_1 tends to produce a more continuous background (as expected) but at the cost of a slight change in strain level — the olive appears less stiff (lighter) in Fig. 4(g). NPR with C_2 is better at preserving absolute strain levels in this case.

In Fig. 5, neither PLLSR nor Gaussian filtering cope well with the vertical band of zero strain (and hence very low precision) down the centre of the image. In both cases, to differing extents, a sharp vertical feature is still visible in the filtered data, even with heavy filtering. Both forms of NPR effectively ignore the data at the centre of the image, and smooth over this region. NPR with C_1 in Fig. 5(f) and (g) works particularly well in this case; even with light filtering, the noise in the central region is very well suppressed.

Fig. 6 contains higher quality raw data and in this case all filtering algorithms produce reasonable strain images, particularly with light filtering. Differences can still be seen in the small noisy regions at the bottom left and underneath the inclusion — these are better suppressed in Fig. 6(f) and (h) than in Fig. 6(b) and (d). However, at heavy filtering there is a noticeable variation in the size of the stiff inclusion. PLLSR in Fig. 6(c) makes the inclusion appear significantly smaller, whilst Gaussian filtering in (e) makes it appear significantly larger. In contrast, in both NPR methods, the inclusion remains much the same size, if anything NPR with C_1 tending towards the inner edge of the low precision white band around the inclusion, and NPR with C_2 tending towards the outer edge.

The phantom data in Figs. 4, 5 and 6 was also used to test convergence for various methods of solving eq. (4). Six iterative methods were investigated:

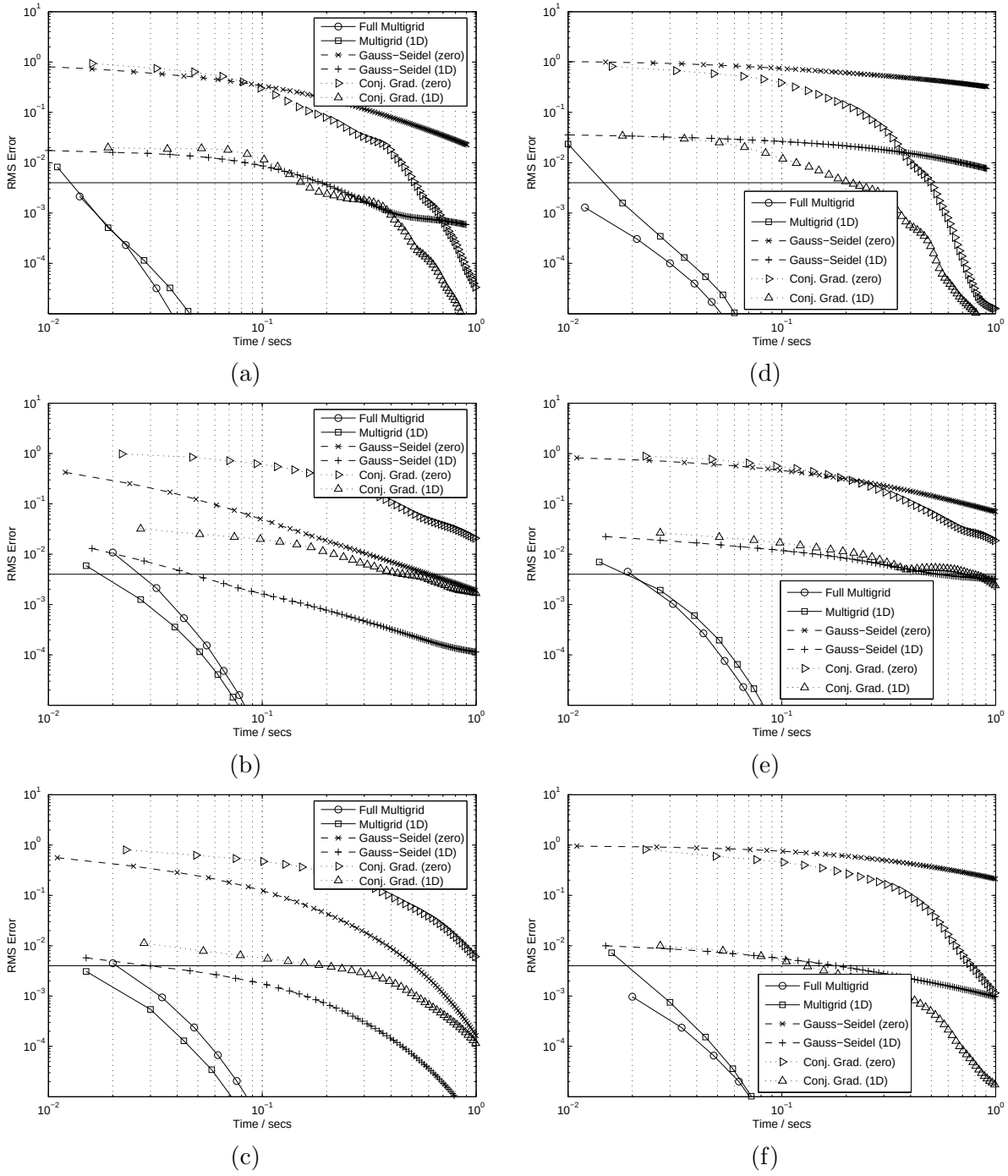


Figure 7: Convergence for C_1 NPR algorithms. Convergence times are shown, on a log-log scale, for various solution strategies with different initial guesses. The left column is for light smoothing, the right column for heavy smoothing. (a) and (d) correspond to Fig. 4(f) and (g) respectively, (b) and (e) correspond to Fig. 5(f) and (g), and (c) and (f) correspond to Fig. 6(f) and (g). The horizontal line shows the point at which further improvement is no longer visible in the strain images. The multigrid solution far outperforms the alternatives, though the difference between full multigrid and multigrid with a pseudo-1D initial guess is only slight.

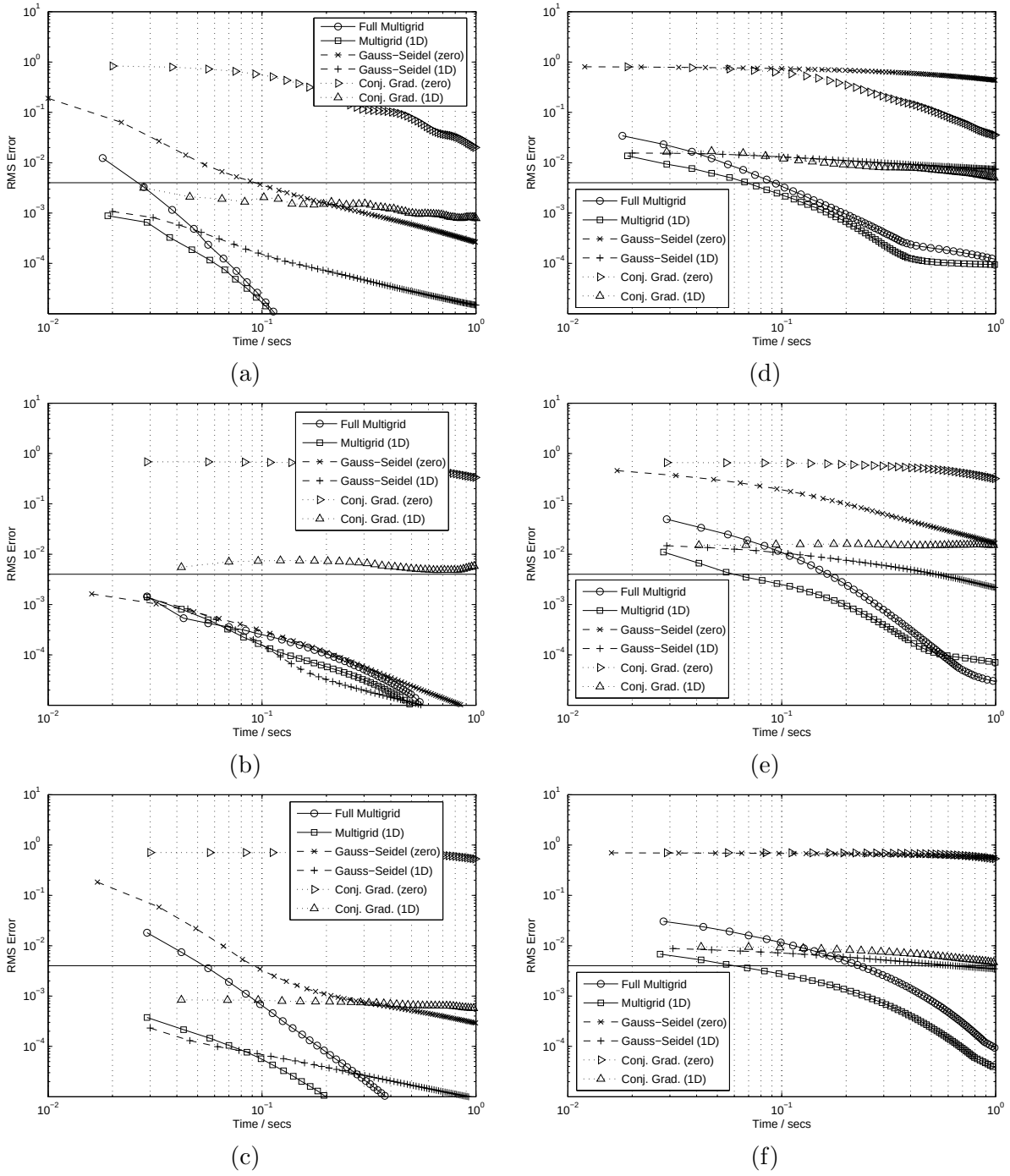


Figure 8: Convergence for C_2 NPR algorithms. Convergence times are shown, on a log-log scale, for various solution strategies with different initial guesses. The left column is for light smoothing, the right column for heavy smoothing. (a) and (d) correspond to Fig. 4(h) and (i) respectively, (b) and (e) correspond to Fig. 5(h) and (i), and (c) and (f) correspond to Fig. 6(h) and (i). The horizontal line shows the point at which further improvement is no longer visible in the strain images. The multigrid solutions far outperform the alternatives, with the pseudo-1D initial guess consistently the best performer.

- **Full Multigrid** This is an iterative scheme which starts by solving the problem directly on a very coarse grid, and hence does not require initialisation, see Appendix B.
- **Multigrid (1D)** Solution using the multigrid framework, and the pseudo-1D initialisation in eq. 5, see Appendix B.
- **Gauss-Seidel (zero)** Red-Black Gauss-Seidel iterations (Varga, 2000), with a zero initial guess.
- **Gauss-Seidel (1D)** Red-Black Gauss-Seidel iterations, and the pseudo-1D initialisation in eq. 5.
- **Conj. Grad. (zero)** The conjugate gradient algorithm (Press et al., 2002), with no preconditioning[9], and a zero initial guess.
- **Conj. Grad. (1D)** The conjugate gradient algorithm, and the pseudo-1D initialisation in eq. 5.

Convergence time was measured on a 2.16 GHz Core2 CPU[10], with the software running single-threaded. For comparison, the time for a direct solution using Cholesky decomposition ranged from 15 seconds (for C_1 on the smaller half-olive phantom data set) to 4 minutes (for C_2 on the larger CIRS phantom data sets). For the Multigrid solutions, one point is plotted after each V-cycle. For the other solutions, one point is plotted after each ten iterations.

The RMS Error shown on each of the graphs in Figs 7 and 8 is the root-mean-squared difference between the current and correct solution. The correct solution was calculated by direct Cholesky decomposition, then subsequent multigrid iterations to improve the rounding accuracy of the result, until there was no further improvement in the residual error. The differences these iterations made on the direct solution were imperceptible in the strain images. The error was calculated over the entire image, except for those pixels which had a precision which had been set to zero due to the lack of a displacement match, for instance right at the base of each image. These areas are shown in red in Figs. 4, 5 and 6. A solid horizontal line on each graph corresponds to an error of less than half of one image grey level, assuming 255 levels of grey in the data. This is the approximate error at which further convergence results in no visible change to the strain image.

Figure 7 contains convergence results for NPR using C_1, on all three data sets, with both light (left column) and heavy (right column) smoothing. Both the Full Multigrid and Multigrid (1D) methods perform remarkably well on this data, producing an image visually indistinguishable from the correct solution within only two iterations, or at most 27 ms. This represents a little over 25% of the total time to convert raw RF data to the displayed strain image. There is little to choose between the two: Full Multigrid performs slightly better than Multigrid (1D) with heavy smoothing, and vice-versa for light smoothing. The other methods with the pseudo-1D initialisation perform reasonably well, but can take up to 1 second to converge. The poor performance of the remaining methods serves to demonstrate how important the pseudo-1D initialisation is.

Figure 8 contains results for NPR using C_2. The convergence is slower than in Fig. 7, since we are now solving a higher order problem. The Full Multigrid and Multigrid (1D) methods are still clearly the best, however now Multigrid with the pseudo-1D initialisation is consistently the better performer of the two, producing a visually correct solution within at most 68 ms. At worst,

[9] In practice, the most obvious choice of preconditioning with the main diagonal of $(\mathbf{W} + \mathbf{M}^T\mathbf{M})$ decreases the number of iterations to convergence, but increases the elapsed time.

[10] Intel corporation, http://www.intel.com

this represents nearly 50% of the total time to convert raw RF data to the displayed strain image. The distinction between light and heavy smoothing is more pronounced than with C_1, with light smoothing needing only one iteration whereas heavy smoothing requires several. The pseudo-1D initialisation is particularly good in the case of light smoothing — itself giving a solution which is visually indistinguishable from the correct one. Hence all methods which make use of it are instantly successful. Heavy smoothing is more problematic for both Conj. Grad. and Gauss-Seidel methods where convergence is very slow indeed.

4.3 Clinical examples

Clinical data was recorded using a Terason T3000[11] ultrasound scanner with a 38 mm linear array wideband probe with centre frequency 7.75 MHz. RF data sampled at 40 MHz was streamed via a software driver to the in-house Stradwin[12] application displaying live B-mode and strain images side by side. All images are of 4 cm depth, with one transmit focus roughly half way down the image.

Figures 9 and 10 contain clinical data from examinations of the breast, thyroid and testis. In each case, a typical B-mode image is shown together with a strain image using precision-weighted Gaussian filtering, and NPR using C_1. The strain images are precision-persisted across several frames, then filtering is applied after persistence. A red colour wash is used to suppress areas of low precision in the Gaussian filtered images, or low resolution in the NPR images.

Real clinical data[13] contains greater variation of ultrasound signal strength, significantly less ideal stress distributions (due to anatomical movement not related to probe contact pressure) and is less homogeneous than the phantom data in Section 4.2. It is not at all clear that NPR filtering with C_1, which seeks to impose continuity on the strain data, should be successful on such images. However, the NPR images in Figs. 9 and 10 show clear advantages over the Gaussian filtered images. Borders between regions of different strain in good precision areas are well represented, whilst regions of apparently rapidly changing strain due to low precision measurement are well suppressed.

4.4 3D strain imaging

NPR has been presented in 2D for clarity, but both the matrix equation (4), pseudo-1D initialisation in eq. (5) and multigrid method are easily extended to 3D. In fact, processing time only increases by a factor of $3N_z/2$, where N_z is the number of frames of data in the elevational (out of plane) direction. 3D scans were performed using a RSP 6-12 MHz[14] linear array probe with an integral stepper motor connected to a Dynamic Imaging Diasus ultrasound machine, with RF acquisition and motor control coordinated with Stradwin software. 3D strain processing details are as described in our previous work (Treece et al., 2008).

Figure 11 shows a 3D scan of the half-olive phantom described in Section 4.2. The strain data contained $127 \times 345 \times 120$ samples. NPR processing, using a mixture of C_1 and C_2[15], took 6.7 seconds,

[11] Terason Ultrasound, Massachusetts, USA, http://www.terason.com.

[12] Available free from http://mi.eng.cam.ac.uk/~rwp/stradwin.

[13] This clinical data is also available in a database currently populated with more than two hundred files acquired through ongoing studies into 2D and 3D freehand quasi-static strain imaging at Addenbrooke's Hospital (Cambridge, UK). Ethical approval was granted by the Cambridgeshire 3 Research Ethics Committee for the anonymized data to be disseminated to the wider research community. Research access to this database can be arranged by emailing stradwin-info@eng.cam.ac.uk. Patients are made aware of this at the time of obtaining informed consent. Anatomical sites currently available include head and neck, breast, testis, kidney, liver and uterus.

[14] GE Healthcare Bio-sciences AB, Uppsala, Sweden, http://www.gehealthcare.com.

[15] This is easily achieved by using $\mathbf{M} = r_1\mathbf{M_1} + r_2\mathbf{M_2}$, where both types of smoothing were given equal precision by setting $r_2 = r_1^2$. In this case we cannot guarantee precisely uniform precision, since $\mathbf{W}$ needs to be set to a different

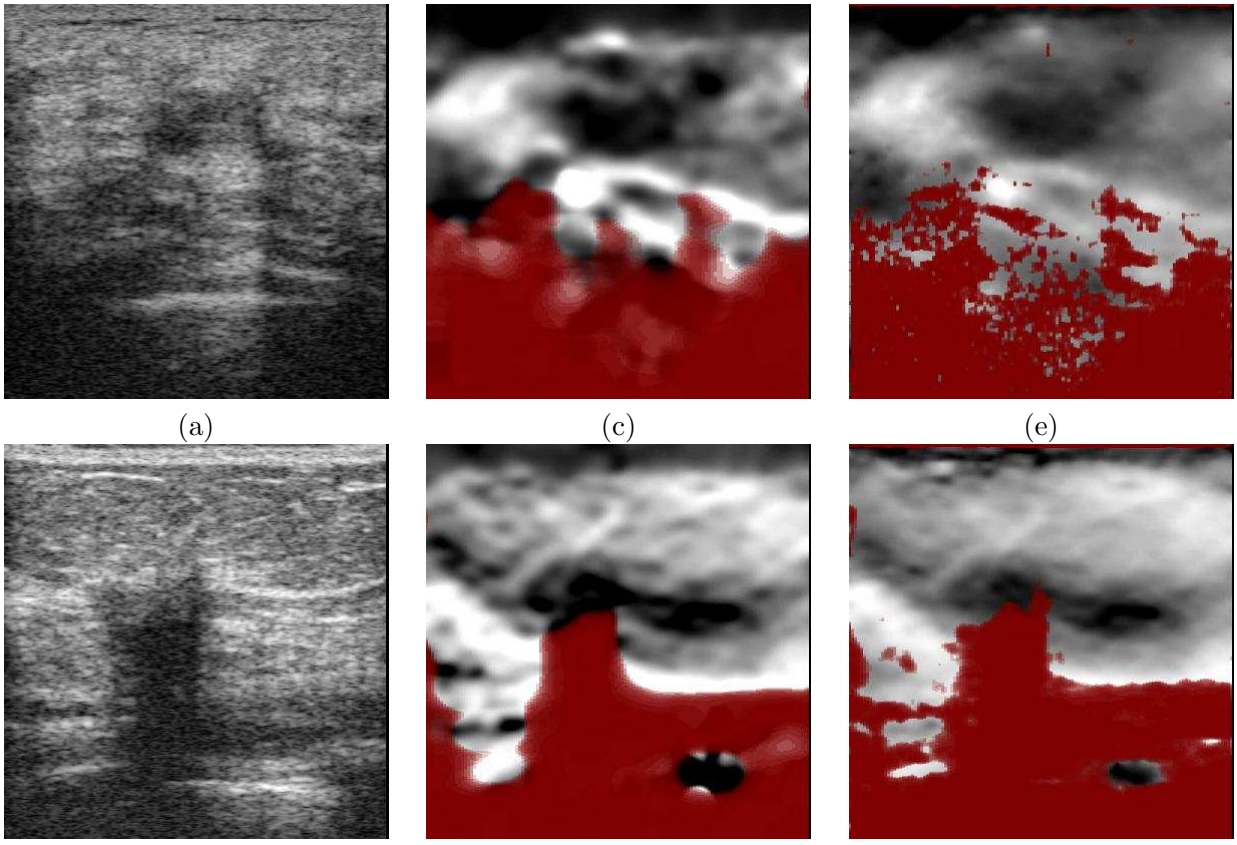


Figure 9: Clinical breast images. (a) B-mode image of a breast fibroadenoma, (c) strain image using Gaussian filtering and (e) using C_1 NPR. (b) B-mode image of a breast invasive carcinoma with surrounding ductal carcinoma in situ, (d) strain image using Gaussian filtering and (f) using C_1 NPR. In both cases the NPR strain images show the mass better delineated, whilst also suppressing distracting strain noise in the remainder of the image. For Gaussian filtering the red colour wash indicates lack of data precision, whereas with NPR it indicates lack of resolution.

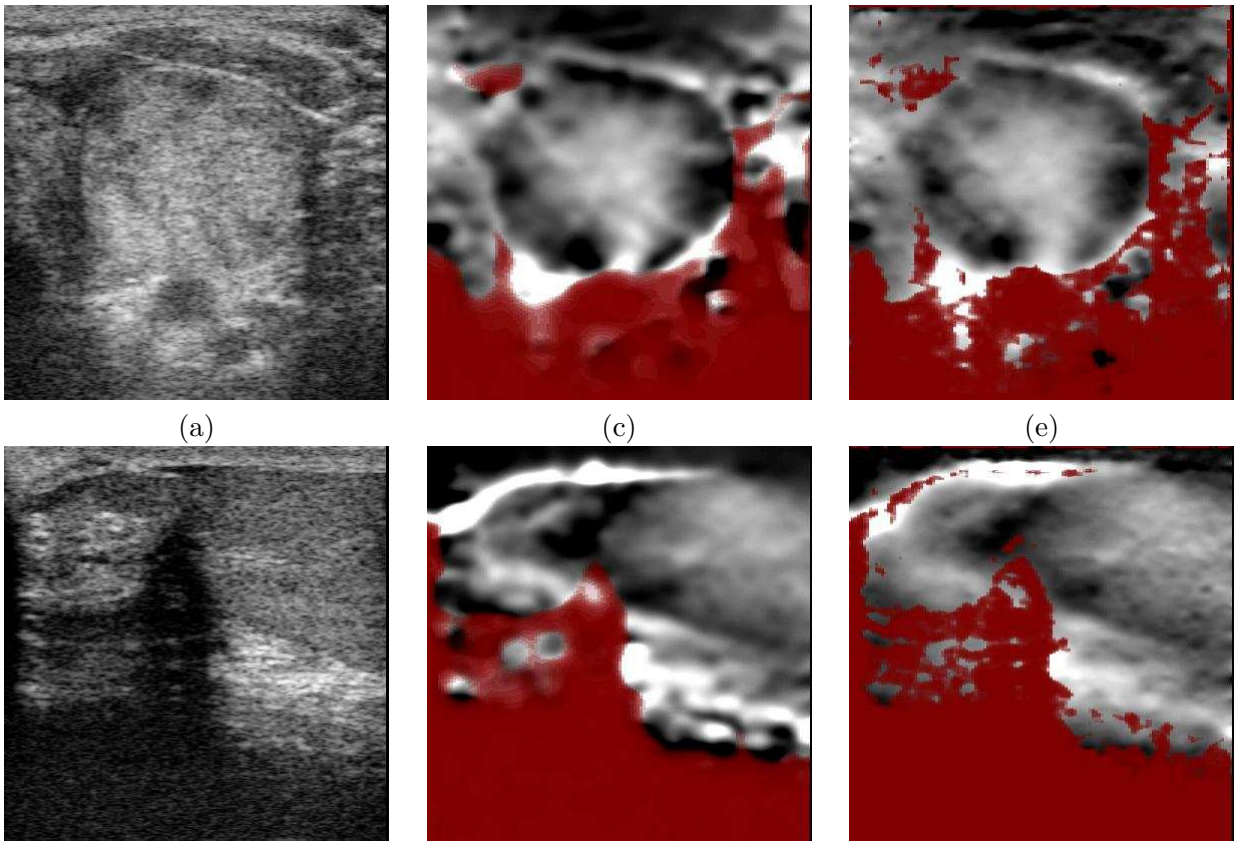
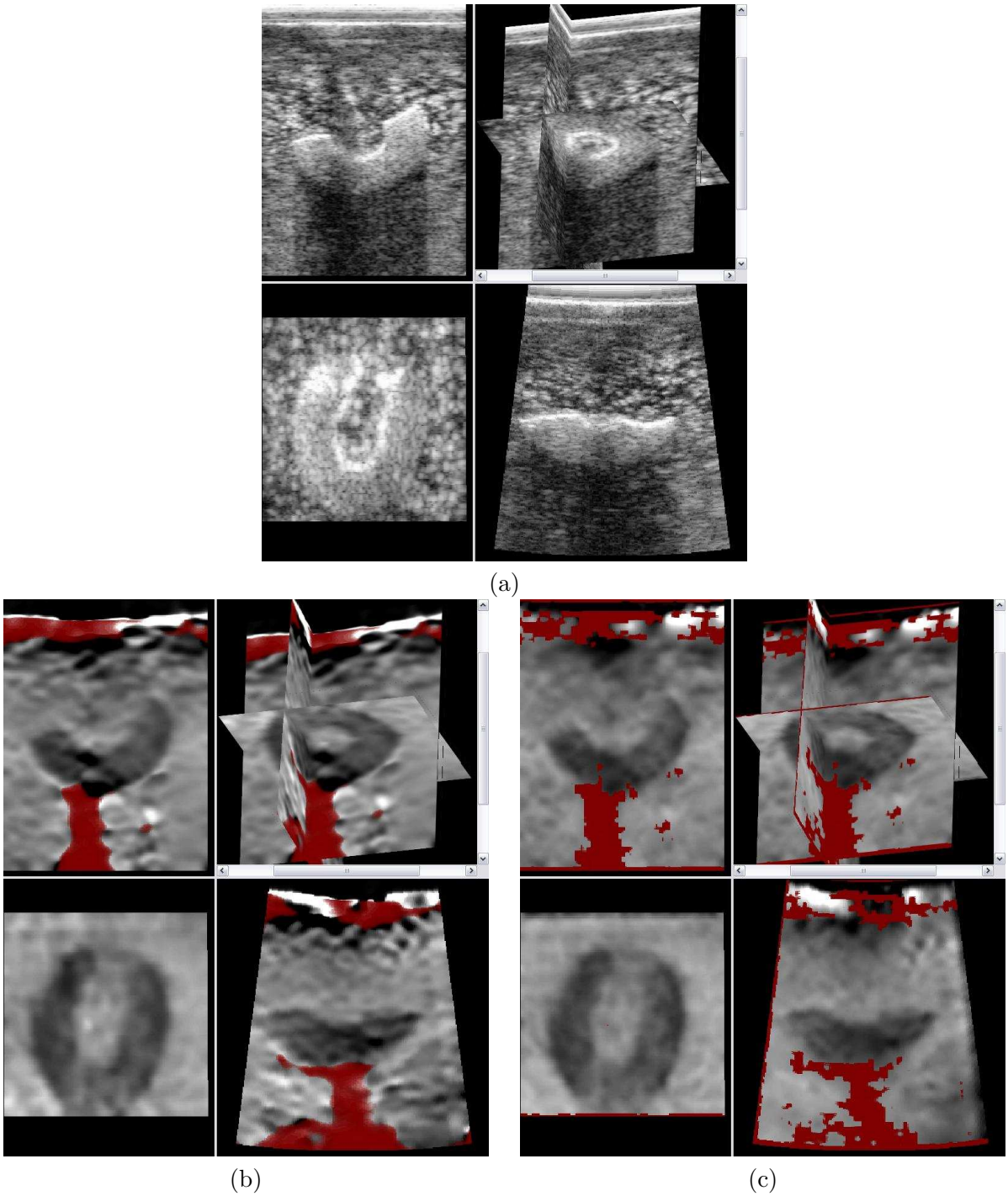


Figure 10: Clinical thyroid and testicular images. (a) B-mode image of a thyroid benign follicular adenoma, (c) strain image using Gaussian filtering and (e) using C_1 NPR. (b) B-mode image showing scrotal calcification, (d) strain image using Gaussian filtering and (f) using C_1 NPR. The improvement due to NPR is similar to Fig. 9. For Gaussian filtering the red colour wash indicates lack of data precision, whereas with NPR it indicates lack of resolution.

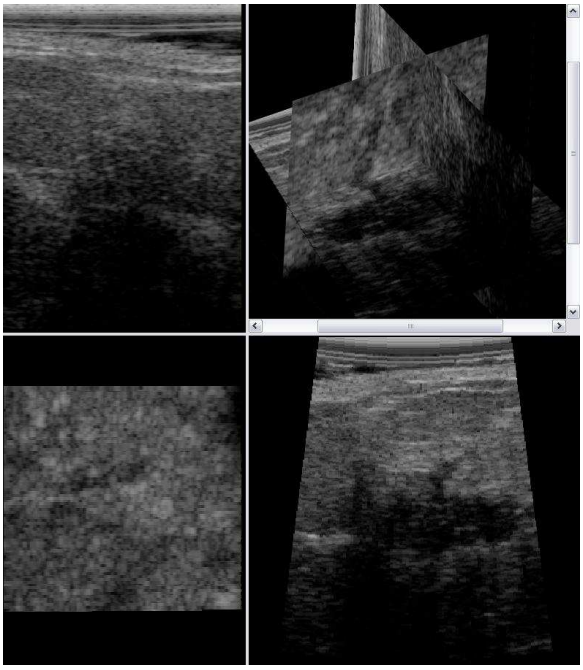


(a)

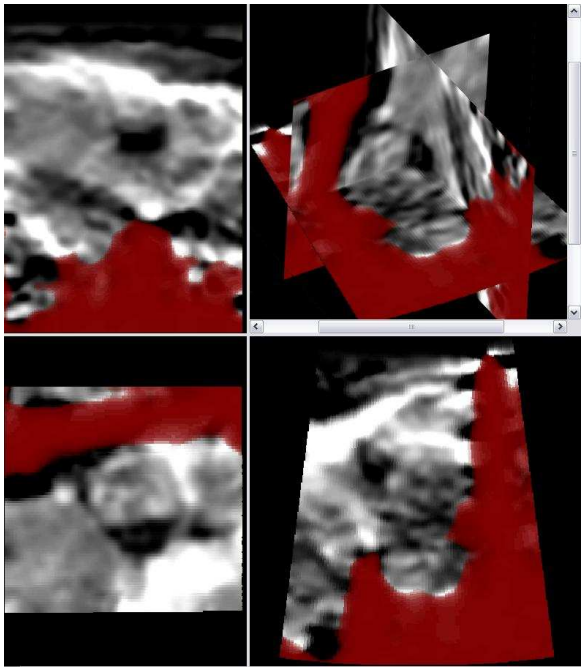
(b)

(c)

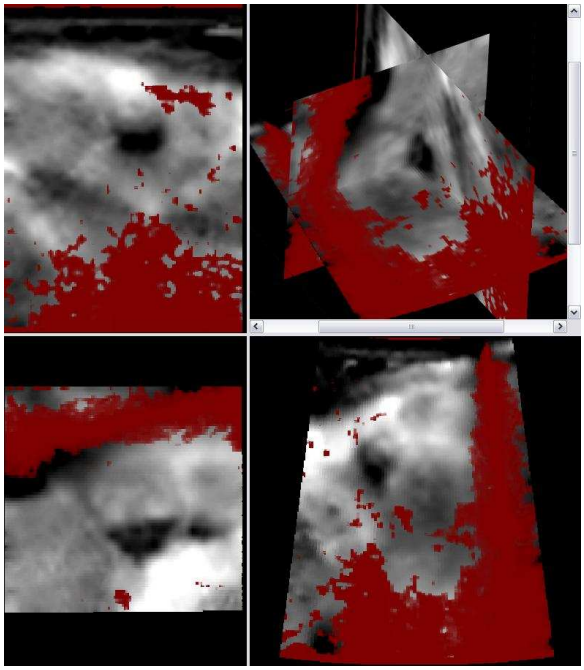
Figure 11: 3D data from the half olive phantom. Each image shows three orthogonal views and a combined 3D view. (a) B-mode data, (b) strain data using Gaussian filtering and (c) using C_1 and C_2 NPR. For Gaussian filtering the red colour wash indicates lack of data precision, whereas with NPR it indicates lack of resolution.



(a)



(b)



(c)

Figure 12: 3D data from a clinical breast examination. Each image shows three orthogonal views and a combined 3D view. (a) B-mode data for a possible invasive ductal carcinoma, (b) strain data using Gaussian filtering and (c) using C_1 and C_2 NPR. For Gaussian filtering the red colour wash indicates lack of data precision, whereas with NPR it indicates lack of resolution.

out of a total 26.8 seconds for generating 3D strain data from the raw RF. Precision-weighted Gaussian filtering took 3.2 seconds out of a total of 23.2 seconds. The NPR result is clearer, particularly towards the top of the data, and in the shadow region under the olive.

Figure 12 is a clinical 3D scan from a breast examination. Here the patient was scanned with 2D strain imaging first to establish the general area of interest before a 3D scan was performed. The strain data contained $127 \times 352 \times 80$ samples. NPR was again applied with a mixture of C_1 and C_2, in this case taking only 3.8 seconds out of a total of 20.4 seconds. Precision-weighted Gaussian filtering took 2.2 seconds out of a total of 18.7 seconds. Two low strain (black) regions are very obvious in the strain images, in an area unrelated to the larger black region in the B-mode image (which was what initially drew the clinicians attention). However, the presentation is confused in the Gaussian-filtered image of Fig. 11(b) by the existence of a large fine scale pattern of strain variation apparent in most of the views. The NPR image in Fig. 11(c) draws attention much more clearly to the real low strain region.

5 Conclusions

We have presented a class of algorithms which can, under certain assumptions, generate strain images with uniform precision but varying resolution. Even when these assumptions are violated, NPR produces images which, though not of exactly uniform precision, still have useful properties: this is demonstrated by the phantom and clinical data sets. Local variation in smoothing is automated, but a single meaningful parameter remains, which directly sets the precision of the filtered data, assuming this is greater than the precision of the raw data. We have developed two specific examples which seek either to impose continuity or smoothness on the strain data, however in practice there may well be other constraints which better model the expected strain behaviour and could be placed in the same framework.

We have demonstrated that NPR using both the C_1 and C_2 constraints can easily be performed in real time on standard computer hardware, by using the multigrid framework with an initial guess at the solution based on a pseudo-1D algorithm. There are some advantages and disadvantages of both constraints. C_1 reliably converges to the correct solution in less than 30 ms in even the fairly large data sets we have tested, however it can modify the strain level if heavy smoothing is applied. C_2 is better at preserving absolute strain levels, but leaves homogeneous regions slightly less constant, and can take longer to converge. However, the efficiency of both algorithms is such that they can still be applied in only a few seconds to fairly large 3D data sets.

Strain images filtered with NPR have a promising appearance when compared to PLLSR or precision-weighted Gaussian filtering, with well defined borders between different strain regions, and good suppression of noise. In addition, NPR seems to preserve the size of stiff objects better than the alternatives tested, which could be important if assessment of size between strain and B-mode images is a clinical factor. Initial clinical examples indicate that the theory still produces useful images even on real data which is significantly more noisy and less homogeneous. However, further clinical studies will be necessary to properly assess the benefits of uniform precision as opposed to uniform resolution strain imaging.

function of the precision in each case, so we set $\mathbf{W}$ for C_1 smoothing.

Acknowledgments

Much of this work was performed while Graham Treece was supported by a Research Fellowship from the Royal Academy of Engineering / EPSRC. He is now partly supported by the Evelyn Trust. Thanks to Luke Sonoka and Sue Freeman for acquiring, and Lujie Chen for organising, the clinical data.

References

Alam, S. K., Ophir, J., Konofagou, E. E., March 1998. An adaptive strain estimator for elastography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 45 (2), 461–472.

Barr, R. G., October 2006. Clinical applications of a real time elastography technique in breast imaging. In: Ultrasonic Measurement and Imaging of Tissue Elasticity. Snowbird, Utah, p. 112.

Briggs, W. L., Henson, V. E., McCormick, S. F., 2000. A multigrid tutorial, 2nd Edition. SIAM.

Catmull, E., Rom, R., 1974. A class of local interpolating splines. In: Barnhill, R., Risenfeld, R. (Eds.), Computer Aided Geometric Design. Academic Press, San Francisco, pp. 317–326.

Céspedes, I., Huang, Y., Ophir, J., Spratt, S., April 1995. Methods for estimation of subsample time delays of digitized echo signals. Ultrasonic Imaging 17 (2), 142–171.

Céspedes, I., Ophir, J., April 1993. Reduction of image noise in elastography. Ultrasonic Imaging 15 (2), 89–102.

Chen, L., Treece, G. M., Lindop, J. E., Gee, A. H., Prager, R. W., 2009. A quality-guided displacement tracking algorithm for ultrasonic elasticity imaging. To appear in Medical Image Analysis .

de Korte, C. L., Céspedes, E. I., van der Steen, A. F. W., Pasterkamp, G., Bom, N., 1998. Intravascular ultrasound elastography: assessment and imaging of elastic properties of diseased arteries and vulnerable plaque. European Journal of Ultrasound 7, 219–224.

de Korte, C. L., Pasterkamp, G., van der Steen, A. F. W., Woutman, H. A., Bom, N., 2000. Characterization of plaque components with intravascular ultrasound elastography in human femoral and coronary arteries *in vitro*. Circulation 102, 617–623.

Emelianov, S. Y., Chen, X., Knipp, B., Myers, D., June 2002. Triplex ultrasound: elasticity imaging to age deep venous thrombosis. Ultrasound in Medicine and Biology 28 (6), 757–767.

Garra, B. S., Céspedes, E. I., Ophir, J., Spratt, S. R., Zuurbier, R. A., Magnant, C. M., Pennanen, M. F., January 1997. Elastography of breast lesions: initial clinical results. Radiology 202 (1), 79–86.

George, A., Liu, J. W., 1981. Computer solution of large sparse positive definite systems. Prentice-Hall.

Green, P. J., Silverman, B. W., 2004. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. CRC Press.

Jensen, J. A., 1996. Field: a program for simulating ultrasound systems. In: Proceedings of the 10th Nordic-Baltic Conference on Biomedical Imaging. Vol. 4. pp. 351–353.

Jiang, J., Hall, T. J., Sommer, A. M., April 2007. A novel image formation method for ultrasonic strain imaging. Ultrasound in Medicine and Biology 33 (4), 643–652.

Kallel, F., Ophir, J., July 1997. A least-squares strain estimator for elastography. Ultrasonic Imaging 19 (3), 195–208.

Kaluzynski, K., Chen, X., Emelianov, S. Y., Skovoroda, S. R., O'Donnell, M., July 2001. Strain rate imaging using two-dimensional speckle tracking. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 48 (4), 1111–1123.

Lindop, J. E., Treece, G. M., Gee, A. H., Prager, R. W., September 2007. Estimation of displacement location for enhanced strain imaging. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 54 (9), 1751–1771.

Lindop, J. E., Treece, G. M., Gee, A. H., Prager, R. W., May 2008a. Dynamic resolution selection in ultrasonic strain imaging. Ultrasound in Medicine and Biology 34 (5), 809–823.

Lindop, J. E., Treece, G. M., Gee, A. H., Prager, R. W., Nov. 2008b. The general properties including accuracy and resolution of linear-filtering methods for strain estimation. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 55 (11), 2362–2368.

Lindop, J. E., Treece, G. M., Gee, A. H., Prager, R. W., Jul. 2008c. An intelligent interface for freehand strain imaging. Ultrasound in Medicine and Biology 34 (7), 1117–1128.

Lindop, J. E., Treece, G. M., Gee, A. H., Prager, R. W., Mar. 2008d. Nonparametric surface regression for strain estimation. Tech. Rep. CUED/F-INFENG/TR 598, Cambridge University Department of Engineering.

Lindop, J. E., Treece, G. M., Gee, A. H., Prager, R. W., Jan. 2008e. Phase-based ultrasonic deformation estimation. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 55 (1), 94–111.

Lubinski, M. A., Emelianov, S. Y., O'Donnell, M., January 1999. Speckle tracking methods for ultrasonic elasticity imaging using short-time correlation. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 46 (1), 82–96.

Maurice, R. L., Bertrand, M., July 1999. Lagrangian speckle model and tissue-motion estimation — theory. IEEE Transactions on Medical Imaging 18 (7), 593–603.

O'Donnell, M., Skovoroda, A. R., Shapo, B. M., Emelianov, S. Y., May 1994. Internal displacement and strain imaging using ultrasonic speckle tracking. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 41 (3), 314–325.

Ophir, J., Céspedes, I., Ponnekanti, H., Yazdi, Y., Li, X., April 1991. Elastography: a quantitative method for imaging the elasticity of biological tissues. Ultrasonic Imaging 13 (2), 111–134.

Pesavento, A., Perrey, C., Krueger, M., Ermert, H., September 1999. A time efficient and accurate strain estimation concept for ultrasonic elastography using iterative phase zero estimation. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 46 (5), 1057–1067.

Pinton, G. F., Dahl, J. J., Trahey, G. E., June 2006. Rapid tracking of small displacements with ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 53 (6), 1103–1117.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 2002. Numerical recipes in C: The art of scientific computing, 2nd Edition. Cambridge University Press.

Regner, D. M., Hesley, G. K., Hangiandreou, N. J., Morton, M. J., Nordland, M. R., Meixner, D. D., Hall, T. J., Farrell, M. A., Mandrekar, J. N., Harmsen, W. S., Charboneau, J. W., February 2006. Breast lesions: evaluation with US strain imaging — clinical experience of multiple observers. Radiology 238 (2), 425–437.

Reinsch, C. H., 1967. Smoothing by spline functions. Numerische Mathematik 10, 177–183.

Silverman, B. W., 1984. Spline smoothing: the equivalent variable kernel method. The Annals of Statistics 12 (3), 898–916.

Srinivasan, S., Ophir, J., Alam, S. K., 2002. Elastographic imaging using staggered strain estimates. Ultrasonic Imaging 24, 229–245.

Sumi, C., 1999. Fine elasticity imaging utilizing the iterative RF-echo phase matching method. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 46 (1), 158–166.

Svensson, W. E., Amiras, D., 2006. Ultrasound elasticity imaging. Breast Cancer Online 9 (6), 1–7.

Treece, G. M., Lindop, J. E., Gee, A. H., Prager, R. W., Oct. 2006. Efficient elimination of dropouts in displacement tracking. In: Proceedings of Ultrasonic Measurement and Imaging of Tissue Elasticity. Snowbird, Utah, USA, p. 68.

Treece, G. M., Lindop, J. E., Gee, A. H., Prager, R. W., 2008. Freehand ultrasound elastography with a 3D probe. Ultrasound in Medicine and Biology 34 (3), 463–474.

Varga, R. S., 2000. Matrix Iterative Analysis. Springer.

Viola, F., Walker, W. F., April 2003. A comparison of the performance of time-delay estimators in medical ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 50 (4), 392–401.

Vogt, M., Ermert, H., March 2005. Development and evaluation of a high-frequency ultrasound-based system for *in vivo* strain imaging of the skin. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 52 (3), 375–385.

Whittaker, E., 1923. On a new method of graduation. Proceedings of the Edinburgh Mathematical Society 41, 63–75.

Zhu, Y., Hall, T. J., July 2002. A modified block matching method for real-time freehand strain imaging. Ultrasonic Imaging 24 (3), 161–176.

Appendices

A Matrices

2D strain data on the regression surface are expressed as a vector of length $N_x N_y$,

$$\acute{\mathbf{s}}^T = \begin{bmatrix} \acute{s}_{(1,1)} & \acute{s}_{(1,2)} & \cdots & \acute{s}_{(1,N_y)} & \acute{s}_{(2,1)} & \cdots & \acute{s}_{(N_x,N_y)} \end{bmatrix}.$$

$\mathbf{M_1}$ and $\mathbf{M_2}$ are $2(N_x N_y - N_x - N_y) \times N_x N_y$ matrices defined such that the vector $\mathbf{M}\acute{\mathbf{s}}$ lists first and second differences respectively, throughout the grid of strain data. For example

$$\mathbf{M_2}\acute{\mathbf{s}} = \begin{bmatrix} (2\acute{s}_{(2,1)} - \acute{s}_{(3,1)} - \acute{s}_{(1,1)}) \\ \vdots \\ (2\acute{s}_{(N_x-1,N_y)} - \acute{s}_{(N_x,N_y)} - \acute{s}_{(N_x-2,N_y)}) \\ (2\acute{s}_{(1,3)} - \acute{s}_{(1,4)} - \acute{s}_{(1,2)}) \\ \vdots \\ (2\acute{s}_{(N_x,N_y-1)} - \acute{s}_{(N_x,N_y)} - \acute{s}_{(N_x,N_y-2)}) \end{bmatrix}$$

Solving for the optimal regression surface involves the inversion of eq. (4). For C_2, the matrix $\mathbf{W} + r\mathbf{M_2}^T\mathbf{M_2}$ is a key part of this expression. Owing to its sparsity, matrix multiplication is never applied directly. Simultaneous equations arising from this matrix associated with coordinates (x, y), away from the edge of the regression surface, have the form

$$\begin{aligned} &\left(w_{(x,y)} + 12r\right)\left[\acute{s}_{(x,y)}\right] \\ &\quad -4r\left[\acute{s}_{(x+1,y)} + \acute{s}_{(x-1,y)} + \acute{s}_{(x,y+1)} + \acute{s}_{(x,y-1)}\right] \\ &\quad +r\left[\acute{s}_{(x+2,y)} + \acute{s}_{(x-2,y)} + \acute{s}_{(x,y+2)} + \acute{s}_{(x,y-2)}\right] \\ &\quad = w_{(x,y)}\hat{s}_{(x,y)} \end{aligned}$$

B Multigrid implementation

Multigrid describes a framework for solving inverse problems rather than a black-box solution — the details of this framework vary with each application. No attempt is made to explain the multigrid or full multigrid framework here, but sufficient application-specific details are given to implement multigrid in this case given general knowledge of the framework (Briggs et al., 2000; Press et al., 2002).

Crucial to the technique are *restriction* and *prolongation* operators, for transferring the residual error to and from coarser grids. For NPR using C_1, the equations are only second order, and in this case we can use simple bi-linear interpolation for prolongation, and its adjoint *full weighting* for restriction. These are described by stencils. For prolongation, the stencil shows the result in the fine grid of prolongating a digital impulse in the coarser grid. For restriction, the stencil shows the weightings applied to the fine grid to generate a data point on the coarse grid. For coarse grids which are exactly half the resolution of the finer grid, a symmetric restriction stencil should be set to $\frac{1}{4}$ of the prolongation stencil for 2D processing, $\frac{1}{8}$ for 3D processing. The prolongation stencil for bi-linear interpolation in 2D is:

$$\frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{9}$$

NPR using C_2 requires the solution of a fourth order equation, and for this we need smoother prolongation and restriction operators. We base these on the Catmull-Rom spline (Catmull and Rom, 1974), which has the useful property in this case of ensuring, for prolongation, that the second order difference of intermediate samples is the exact bi-linear interpolation of the second order difference of the surrounding samples in the coarser grid. In 1D, the intermediate data point is given by applying the following weightings to the coarser grid data:

$$\frac{1}{16}\begin{bmatrix} -1 & 9 & 9 & -1 \end{bmatrix} \tag{10}$$

The full prolongation stencil for 2D is hence

$$\frac{1}{256}\begin{bmatrix} 1 & 0 & -9 & -16 & -9 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -9 & 0 & 81 & 146 & 81 & 0 & -9 \\ -16 & 0 & 146 & 256 & 146 & 0 & -16 \\ -9 & 0 & 81 & 146 & 81 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -9 & -16 & -9 & 0 & 1 \end{bmatrix} \tag{11}$$

The same restriction operators were used to restrict the data weights $\mathbf{W}$ to each grid level. These operations can easily be implemented as 1D convolutions in each of the x and y directions.

Multiple V-cycles were used, with Red-black Gauss-Seidel (Varga, 2000) as the smoothing operator on each grid, except for the coarsest grid (of at least 6×6), where a direct solution was found using band-limited Cholesky decomposition. For NPR using C_1, there were 4 iterations of smoothing at each smoothing stage of the V-cycle. For NPR using C_2, with heavy smoothing, a larger number of smoothing iterations were potentially required at each stage to ensure rounding errors were sufficiently damped. Hence the iteration count was initialised to 4, then increased by 4 whenever the residual error decreased by less than 1% over a whole V-cycle.

C Weighted phase variance

It was shown by Lindop et al. (2008a) that, under certain simplifying assumptions, the precision p of the displacement data can be estimated from the complex cross-correlation ρ of matched displacement windows in the pre- and post-deformation data. For two matched signals $r_1 e^{i\theta_1}$ and $r_2 e^{i\theta_2}$

$$p \approx \frac{\rho}{1-\rho} \tag{12}$$

$$\text{where} \quad \rho = \Re\left[\frac{\sum r_1 r_2 e^{i(\theta_1-\theta_2)}}{\sqrt{\sum r_1^2 \sum r_2^2}}\right] \tag{13}$$

In practice, this calculation is only valid for reasonably high correlations, and a slight modification is made to ensure that low correlations lead to sufficiently low precision values

$$p_\rho = \begin{cases} \frac{3\rho-2}{1-\rho} & \rho > \frac{2}{3} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

When tracking displacements using Weighted Phase Separation (WPS) (Lindop et al., 2008e), the windows are matched by directly minimising a weighted sum of phase differences between pre-

and post-deformation data. In this case, the weighted variance of the residual phase differences σ_θ between each sample of the matched windows is an alternative measure of the quality of the displacement estimate. If the phase difference is weighted by the product of the envelope of each signal $r_1 r_2$, then

$$\sigma_\theta = \frac{\sum r_1 r_2 \left(\theta_1 - \theta_2\right)^2}{\sum r_1 r_2} \tag{15}$$

However, this can be related to eq. (13) by assuming that the residual phase differences $\theta_1 - \theta_2$ will be small, and re-writing

$$\rho \approx \Re \left[\frac{\sum r_1 r_2 + i \sum r_1 r_2 \left(\theta_1 - \theta_2\right) - \frac{1}{2} \sum r_1 r_2 \left(\theta_1 - \theta_2\right)^2}{\sqrt{\sum r_1^2 \sum r_2^2}} \right] \tag{16}$$

The second term in the numerator of eq. (16) is imaginary, and in any case will by definition be zero for matched windows, hence

$$\rho \approx \frac{\sum r_1 r_2}{\sqrt{\sum r_1^2 \sum r_2^2}} - \frac{\sum r_1 r_2 \left(\theta_1 - \theta_2\right)^2}{2\sqrt{\sum r_1^2 \sum r_2^2}} \tag{17}$$

If we also assume that the signal envelopes r_1 and r_2 are similar, i.e. $\sqrt{\sum r_1^2 \sum r_2^2} \approx \sum r_1 r_2$, then we have

$$\rho \approx 1 - \frac{\sigma_\theta}{2} \tag{18}$$

By substituting this into eq. (14) we can now relate precision to the weighted residual phase variance:

$$p_\theta = \begin{cases} \frac{2 - 3\sigma_\theta}{\sigma_\theta} & \sigma_\theta < \frac{2}{3} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

This is useful since eq. (19) can be calculated more efficiently than eq. (14) when using WPS, and is a more direct measure of the residual error in this form of displacement tracking.