

---

**WAVELET RESTORATION OF MEDICAL  
PULSE-ECHO ULTRASOUND IMAGES  
IN AN EM FRAMEWORK**

J. K. H. Ng, R. W. Prager, N. G. Kingsbury,  
G. M. Treece and A. H. Gee

**CUED/F-INFENG/TR 554**

22 May 2006

University of Cambridge  
Department of Engineering  
Trumpington Street  
Cambridge CB2 1PZ  
United Kingdom

Email: [jkhn2/rwp/ngk/gmt11/ahg@eng.cam.ac.uk](mailto:jkhn2/rwp/ngk/gmt11/ahg@eng.cam.ac.uk)

---



# Wavelet Restoration of Medical Pulse-Echo Ultrasound Images in an EM Framework

James Ng, Richard Prager, Nick Kingsbury, Graham Treece, Andrew Gee

University of Cambridge, Department of Engineering,

Trumpington Street, Cambridge CB2 1PZ

## Abstract

The clinical utility of pulse-echo ultrasound images is severely limited by inherent poor resolution which impacts negatively on their diagnostic potential. Research into the enhancement of image quality has mostly been concentrated in the areas of blind image restoration and speckle removal with little regard for accurate modelling of the underlying tissue reflectivity that is imaged. The acoustic response of soft biological tissues has statistics which differ substantially from the natural images considered in mainstream image processing: while, on a macroscopic scale, the overall tissue echogenicity does behave somewhat like a natural image and varies piecewise smoothly, on a microscopic scale, the tissue reflectivity exhibits a microscopic pseudo-random texture (manifested in the amplitude image as speckle) due to the dense concentrations of small, weakly-scattering particles. Recognising that this pseudo-random texture is diagnostically important for tissue identification, we propose modelling tissue reflectivity as the product of a piecewise smooth echogenicity map and a field of uncorrelated, identically distributed random variables. We demonstrate how this model of tissue reflectivity can be exploited in an expectation-maximisation (EM) algorithm that simultaneously solves the image restoration problem and the speckle removal problem by iteratively alternating between Wiener filtering (to solve for the tissue reflectivity) and wavelet-based denoising (to solve for the echogenicity map). Our simulation and *in vitro* results indicate that our EM algorithm is capable of producing restored images which have better image quality and greater fidelity to the true tissue reflectivity than other restoration techniques based on simpler regularising constraints.

## 1 Introduction

The application of pulse-echo ultrasound to anatomical imaging, in particular to the imaging of soft biological tissues, is now well established in medical diagnostics, but despite the advantages pulse-echo ultrasound enjoys over other imaging modalities in terms of equipment cost and patient safety, its clinical utility is limited by poor image resolution due to the finite temporal bandwidth and the non-negligible dimensions of the pulse-echo acoustic beam relative to the size of the scatterers. There is therefore significant scope for the development of computational algorithms to improve the resolution of pulse-echo ultrasound images, a problem we refer to as *image restoration*.

This problem is non-trivial because the blurring effect of the imaging system causes a loss of information in the acquired image which has to be compensated for by the incorporation of prior knowledge based on models of the tissue reflectivity. Successful image restoration requires accurate models of the underlying tissue reflectivity which can adequately capture significant image features without being computationally prohibitive. However, the question of developing appropriate models for tissue reflectivity has been largely neglected, and while there are several papers (e.g. [1, 2, 3, 4]) which address in detail the problem of identifying the blurring operator directly from pulse-echo ultrasound data (the so-called *blind restoration* problem), very few have addressed the

more fundamental problem of accurate yet simple modelling of the tissue reflectivity. Therefore, in this paper, we focus on the development of such a model for the tissue reflectivity and we show how it can be exploited in an expectation-maximisation (EM) algorithm to yield a computationally efficient iterative restoration method based on Wiener filtering and wavelet processing, with the implicit assumption that the blurring operator is known *a priori*. In practice, variations in the speed of sound in the medium can lead to aberrations which distort the blurring operator, but for the purposes of this work, we assume that such effects are either absent or have been corrected for (see, for example, [5, 6, 7]).

Wavelets have been very successful in the processing of so-called *natural images*, i.e. images which are piecewise smooth, because they provide a linear basis in which natural images are sparsely represented and in which their statistical dependencies are substantially simplified. This sparsification and simplification of statistical dependencies has led to the development of simple but computationally efficient wavelet-based algorithms for image compression and analysis (see, for example, [8, 9]). Many of the algorithms for image restoration in the mainstream image processing literature, however, cannot be directly applied to pulse-echo ultrasound images because of fundamental differences in image statistics. The typical reflectivities of soft biological tissues which are imaged by pulse-echo ultrasound do not exhibit piecewise smoothness in the same way as the natural images considered in mainstream image processing: while, on a macroscopic level, the echogenicities of soft tissues appear piecewise smooth and resemble natural images, on a microscopic level, the reflectivities exhibit a pseudo-random behaviour which manifests itself as the characteristic speckle pattern seen in ultrasound images.

Speckle obscures significant image features and degrades the resolvability of structures, and to address this problem, several methods for the removal of speckle have been published in the technical literature (for some of the most recent, see [10, 11, 12, 13, 14]). However, despite its negative effect on image quality, speckle also contains useful textural information that can assist in identification of tissue type. Hence, we believe that restoration methods for pulse-echo ultrasound aimed at improving image quality for diagnostic purposes are required to not only enhance image resolution but to also preserve the textural information inherently present in speckle. Nevertheless, we recognise that speckle removal is useful for tasks that examine gross image features, such as segmentation.

We approach the two-fold problem of enhancing resolvability and preserving textural information by modelling tissue reflectivity as the product of a piecewise smooth component and a random component to account for the macroscopic variations in echogenicity and the microscopic pseudo-random detail respectively (see Fig. 1). Splitting the tissue reflectivity in this way into two distinct components allows us to model their different behaviours separately; furthermore, we observe that, since the echogenicity component is piecewise smooth, we can exploit wavelet-domain models for natural images to describe it. We shall show that, by applying EM, we are able to develop an iterative restoration algorithm that successively alternates between Wiener filtering to estimate the tissue reflectivity and wavelet-domain denoising to estimate the echogenicity. We point out that, because it also produces an estimate of the echogenicity (which, by definition, is free of speckle), our algorithm effectively solves both the image restoration problem and the speckle removal problem simultaneously.

The idea of modelling the echogenicity and the pseudo-random texture separately is not new and was previously proposed by Husby *et al.* [15] who modelled the echogenicity as a Markov random field (MRF) in the image domain and assumed Gaussian statistics for the pseudo-random texture. We have also assumed Gaussian statistics for the pseudo-random texture but have chosen to model the echogenicity in the wavelet domain rather than in the image domain. Another major difference between their approach and ours is that they estimated the echogenicity using Markov chain Monte Carlo (MCMC) methods, whereas our wavelet-domain formulation allows us to estimate the echogenicity with simple non-linear shrinkage rules which are computationally much less intensive. There are also some minor differences: they assumed the blurring operator to be shift-invariant and performed their calculations on real-valued radio-frequency (RF) quantities, whereas we have allowed for a more realistic shift-variant blurring operator and operate on complex-valued baseband quantities which retain the same phase information but can be sampled more sparsely (the sparser sampling, in turn, further reduces the computational load).

We acknowledge the work of Figueiredo and Nowak [16] who developed an iterative algorithm based on EM for the restoration of blurred natural images which also alternates between Wiener filtering and wavelet-domain denoising. Their work, in turn, may be viewed as an extension of Neelamani *et al.*'s [17] Fourier-wavelet regularised deconvolution (ForWaRD) method which applies an under-regularised Wiener filter to the blurred image and follows this up with wavelet denoising. We emphasise, however, that although there is a procedural similarity between our algorithm and these others, the similarity is somewhat superficial and our algorithm was developed for tissue reflectivities with image statistics that are significantly different to those of natural images.

In the sections that follow, we first formulate a model for weak, linear scattering in continuous space-time and derive an equivalent baseband model with complex-valued quantities. We then present the discrete approximation to this continuous space-time model and discuss a method for approximating the global shift-variant blurring operator by a collection of locally shift-invariant blurring operators. We present our model of tissue reflectivity and show how EM can be applied to develop an iterative image restoration algorithm. Finally, we present simulation and *in vitro* results to compare the performance of our algorithm with the performance of more traditional approaches. Mathematical symbols that occur frequently throughout the paper are defined in Table 1.

## 2 Background

### 2.1 Continuous Formulation

In the following description of pulse-echo ultrasound imaging, we use the term *axial* to denote the principal direction of wave propagation, *lateral* to denote the direction perpendicular to this in the plane of the image, and *elevational* to denote the direction perpendicular to the plane of the image.

A conventional pulse-echo ultrasound system acquires an image by transmitting RF pressure pulses of short duration into soft tissue from an array of electromechanical transducers (usually piezoelectric elements) and detecting echoes reflected and backscattered by acoustic inhomogeneities with the same array. Lateral focussing is achieved electronically by coherently exciting elements within a subset of the array (referred to as the *active aperture*) so that the pressure pulses from each excited element converge on the focal point at the same time. Echoes detected within the active aperture are beamformed into a so-called *A-line*. To acquire a complete two-dimensional image, the active aperture is translated laterally to acquire multiple A-lines; since the amplitudes of the A-lines are strongly correlated with structural detail, an image of the anatomy can be viewed by lining up the A-lines in image space and displaying their amplitudes as grey-scale levels. Because these amplitude images usually have very large dynamic range, it is common to display them logarithmically compressed as so-called *B-scan* images.

To describe the imaging system mathematically, we adopt a three-dimensional coordinate system  $(x, y, z)$  in which the  $x$ ,  $y$  and  $z$  axes are aligned with the lateral, elevational and axial directions respectively. We specify the centre of the active aperture to be at  $(x_0, y_0, 0)$  and we use  $t$  to denote time. Under the assumption of linear, weak scattering, the Born approximation allows us to neglect multiple scattering and, at distances sufficiently far away from the active aperture, we may express the A-line  $g(x_0, y_0, t)$  as the Fredholm integral [18]

$$g(x_0, y_0, t) = \iiint h(x_0 - x, y_0 - y, z, t) f(x, y, z) dx dy dz \quad (1)$$

where  $f(x, y, z)$  is the tissue reflectivity and encapsulates spatial variations in density and compressibility and  $h(x, y, z, t)$  is a blurring kernel that accounts for focussing, propagation, scattering and beamforming. All three quantities in (1) are real-valued.

The A-line  $g(x_0, y_0, t)$  is an RF signal, but for subsequent processing, it is more convenient to demodulate it to baseband where it can be sampled at a lower rate without violating the sampling theorem [18]. To

demodulate the A-line, we first suppress its negative frequency components to yield the corresponding complex-valued analytic signal (which may be expressed as  $g(x_0, y_0, t) + j \mathcal{H}_t \{g(x_0, y_0, t)\}$  where  $\mathcal{H}_t \{\bullet\}$  is the temporal Hilbert transform) and we then multiply this analytic signal by  $e^{-j\omega_0 t}$  where  $\omega_0$  is a suitably-defined centre frequency of the A-line. The complex-valued baseband equivalent of (1) may then be written as [18]

$$\tilde{g}(x_0, y_0, t) = \iiint \tilde{h}(x_0 - x, y_0 - y, z, t) \tilde{f}(x, y, z) dx dy dz \quad (2)$$

where

$$\tilde{g}(x_0, y_0, t) = e^{-j\omega_0 t} [g(x_0, y_0, t) + j \mathcal{H}_t \{g(x_0, y_0, t)\}], \quad (3)$$

$$\tilde{h}(x, y, z, t) = e^{-j(\omega_0 t - 2k_0 z)} [h(x, y, z, t) + j \mathcal{H}_t \{h(x, y, z, t)\}], \quad (4)$$

$$\tilde{f}(x, y, z) = e^{-j2k_0 z} f(x, y, z) \quad (5)$$

and  $k_0$  is the wave number, calculated by dividing  $\omega_0$  by the speed of sound in the medium.

If we imagine soft tissue to consist of point scatterers, we may interpret the complex-valued tissue reflectivity  $\tilde{f}(x, y, z)$  defined in (5) as a collection of complex-valued phasors with amplitudes given by  $|f(x, y, z)|$  and phases determined by the axial distance  $z$  from the active aperture (and, strictly speaking, also by the sign of  $f(x, y, z)$ ). In the case of weak scattering, the scatterers are much smaller than the dimensions of the acoustic pulse and their location within the volume of the acoustic pulse is effectively random; hence, the complex phasors in  $\tilde{f}(x, y, z)$  may be considered to have random phase uniformly distributed between 0 and  $2\pi$  [18].

Both (1) and (2) describe the interaction of the acoustic pulse and the insonified subject in three dimensions, but in practice only two-dimensional images are recorded. We may approximate a two-dimensional RF image as the result of blurring a hypothetical two-dimensional tissue reflectivity formed by accumulating the three-dimensional tissue reflectivity along the elevational direction. Bamber and Dickinson [19] showed this approximation to be exact when the blurring kernel is separable: if we let  $\tilde{h}(x, y, z, t) = \tilde{h}'(x_0 - x, z, t) \tilde{h}_0(y, t)$ , define  $\tilde{f}'(x, z) = \int \tilde{h}_0(y_0 - y, t) \tilde{f}(x, y, z) dy$  and let  $\tilde{g}(x_0, y_0, t) = \tilde{g}'(x_0, t)$ , then we may write the two-dimensional analogue of (2) as

$$\tilde{g}'(x_0, t) = \iint \tilde{h}'(x_0 - x, z, t) \tilde{f}'(x, z) dx dz. \quad (6)$$

## 2.2 Discrete Formulation

In practice, processing of the ultrasound image is carried out on digitised data, and so we formulate an equivalent discrete model using matrix-vector notation. We define  $\mathbf{y}$  to be an  $N \times 1$  vector of lexicographically arranged samples of the demodulated RF image,  $\mathbf{x}$  to be a similar  $N \times 1$  vector of complex-valued tissue reflectivity samples,  $H$  to be the complex-valued  $N \times N$  blurring matrix that approximates the continuous space-time blurring operator and  $\mathbf{n}$  to be an  $N \times 1$  noise vector to account for measurement error. The discrete analogue of the Fredholm integral, including the noise vector  $\mathbf{n}$ , is then given by

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}. \quad (7)$$

In this context, the problem of image restoration is to estimate  $\mathbf{x}$  given knowledge of  $\mathbf{y}$ ,  $H$  and the statistics of  $\mathbf{n}$ . We allow  $\mathbf{n}$  to be complex-valued and we assume that it is white, has zero mean and obeys a multivariate complex Gaussian distribution as defined in [20, chapter 15]; defining its covariance matrix  $E(\mathbf{nn}^H) = 2\sigma_n^2 I_N$ , and with a slight abuse of notation, we may write the probability density function of  $\mathbf{n}$  as

$$p(\mathbf{n} | \sigma_n) = \frac{1}{(2\pi\sigma_n^2)^N} \exp\left(-\frac{\|\mathbf{n}\|^2}{2\sigma_n^2}\right) \quad (8)$$

where  $\sigma_n$  is necessarily real-valued. The real and imaginary parts of each element of  $\mathbf{n}$  are uncorrelated and have a variance of  $\sigma_n^2$  each.

In practice, the matrix  $H$  is too large to be explicitly computed and multiplications by  $H$  and  $H^H$  are performed indirectly. In typical applications where the blurring operator is shift-invariant and can be characterised by a single impulse response,  $H$  is block circulant with circulant blocks (BCCB) and can be diagonalised by the discrete Fourier transform (DFT). A product of the form  $H\mathbf{x}$  is then just a discrete convolution of  $\mathbf{x}$  with the impulse response and can be efficiently computed by multiplying together the DFT coefficients of  $\mathbf{x}$  and the DFT coefficients of the impulse response and taking the inverse DFT of the product. Likewise, a product of the form  $H^H\mathbf{x}$  is a discrete convolution of  $\mathbf{x}$  with the complex conjugate of the spatially-reversed impulse response and can be computed by a similar multiplication in the DFT domain with the DFT coefficients of the impulse response replaced by their complex conjugates.

Unfortunately, in pulse-echo ultrasound imaging, the width of the acoustic beam interrogating the subject is non-uniform and varies with axial distance, giving rise to a blurring operator that is shift-variant. Michailovich and Adam [1] proposed handling this shift-variant blurring operator with Nagy and O’Leary’s method of partitioning image space into a number of regions within which the blurring operator is assumed to be locally shift-invariant [21, 22]. By defining an impulse response within each of these approximately shift-invariant regions, we can approximate the global blurring operator by convolving each region with its impulse response, weighting the results and summing them. Mathematically, we may write

$$H \approx \sum_n D_n H_n \quad (9)$$

where  $H_n$  is the BCCB matrix corresponding to the  $n$ th region and  $\{D_n\}$  are diagonal weighting matrices with non-negative elements (we require  $\{D_n\}$  to satisfy  $\sum_n D_n = I_N$  where  $I_N$  is the  $N \times N$  identity matrix). For multiplication by  $H^H$ , we similarly have

$$H^H \approx \sum_n H_n^H D_n \quad (10)$$

which is equivalent to weighting each shift-invariant region, convolving each weighted region with the complex conjugate of its spatially-reversed impulse response and summing the results.

This approximation for the blurring matrix  $H$  is equivalent to approximating the blurring kernel by interpolating between its known values at particular locations in image space. The choice of the weighting matrices  $\{D_n\}$  determines the type of interpolation used. In [21, 22], piecewise constant and piecewise linear interpolation were suggested and, in our experiments, we have chosen to use piecewise linear interpolation.

### 3 Maximum A Posteriori (MAP) Estimation

The problem of image restoration is ill-posed because blurring constitutes a loss of information that is irreversible and so exact recovery of the tissue reflectivity  $\mathbf{x}$  is, in practice, impossible. This is reflected mathematically by the fact that the blurring matrix  $H$  is often singular and cannot be inverted, but even in those cases that it is not singular, it is nevertheless highly ill-conditioned and multiplying  $\mathbf{y}$  by  $H^{-1}$  would amplify the noise term  $\mathbf{n}$  and render the solution physically infeasible (and hence worthless).

The traditional way of coping with the ill-conditioned nature of  $H$  is to impose a regularising constraint which biases our estimate of  $\mathbf{x}$  away from physically infeasible solutions. This usually leads to the optimisation of a cost function which trades off fidelity to the observed image  $\mathbf{y}$  against the regularising constraint on  $\mathbf{x}$ . In a Bayesian setting, this corresponds exactly with maximum *a posteriori* (MAP) estimation: if we treat  $\mathbf{x}$  and  $\mathbf{y}$  as random vectors to which probability density functions are assigned, then MAP estimation seeks the realisation of  $\mathbf{x}$  which maximises its posterior probability  $p(\mathbf{x} | \mathbf{y}, \sigma_n)$ . Bayes’s rule states that  $p(\mathbf{x} | \mathbf{y}, \sigma_n) \propto p(\mathbf{y} | \mathbf{x}, \sigma_n) p(\mathbf{x} | \sigma_n)$ ,

and since  $\mathbf{x}$  and  $\sigma_n$  are independent,  $p(\mathbf{x} | \sigma_n) = p(\mathbf{x})$ . Taking logarithms, our MAP estimate  $\hat{\mathbf{x}}$  may be written as

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} [\ln p(\mathbf{y} | \mathbf{x}, \sigma_n) + \ln p(\mathbf{x})] \quad (11)$$

where the log-likelihood  $\ln p(\mathbf{y} | \mathbf{x}, \sigma_n)$  enforces fidelity to  $\mathbf{y}$  and the log-prior  $\ln p(\mathbf{x})$  is the regularising constraint that reflects our prior belief about  $\mathbf{x}$ . We may substitute  $\mathbf{n} = \mathbf{y} - H\mathbf{x}$  into the probability density function in (8) to give the likelihood distribution

$$p(\mathbf{y} | \mathbf{x}, \sigma_n) = -\frac{1}{(2\pi\sigma_n^2)^N} \exp\left(-\frac{1}{2\sigma_n^2} \|\mathbf{y} - H\mathbf{x}\|^2\right) \quad (12)$$

and, taking logarithms and discarding constants, we may write our MAP estimate as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left[ \frac{1}{2\sigma_n^2} \|\mathbf{y} - H\mathbf{x}\|^2 - \ln p(\mathbf{x}) \right]. \quad (13)$$

### 3.1 $l^2$ -norm Regularisation

One of the more common methods of regularisation is to constrain the weighted  $l^2$ -norm of  $\mathbf{x}$ , a technique known as *Tikhonov regularisation*. Constraining the weighted  $l^2$ -norm of  $\mathbf{x}$  is equivalent to modelling  $\mathbf{x}$  as a Gaussian random vector with zero mean and covariance matrix  $E(\mathbf{x}\mathbf{x}^H) = 2C_x$ ; we may then write  $p(\mathbf{x}) \propto \exp(-\frac{1}{2}\mathbf{x}^H C_x^{-1} \mathbf{x})$  and

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left( \frac{1}{2\sigma_n^2} \|\mathbf{y} - H\mathbf{x}\|^2 + \frac{1}{2} \mathbf{x}^H C_x^{-1} \mathbf{x} \right) = (H^H H + \sigma_n^2 C_x^{-1})^{-1} H^H \mathbf{y}. \quad (14)$$

The closed-form expression on the rightmost side of (14) is the well-known Wiener filter and its application to pulse-echo ultrasound images was proposed by Taxt and Strand [3, 4] and by Michailovich and Adam [1]. The special case  $C_x = \sigma_x^2 I_N$ , where  $\sigma_x$  is real-valued, is referred to as *zero-order Tikhonov regularisation*.

The ForWaRD method, which was briefly mentioned in Section 1, is a non-linear, wavelet-based extension of zero-order Tikhonov regularisation: it applies an under-regularised Wiener filter (formed by replacing  $\sigma_n^2$  with  $\alpha\sigma_n^2$  where  $\alpha \in (0, 1]$  is an under-regularisation parameter) to  $\mathbf{y}$  and then follows this up with wavelet shrinkage to remove the noise amplified by the under-regularisation. The ForWaRD method has been successful in the restoration of natural images and its application to pulse-echo ultrasound images was proposed by Wan *et al.* in [2].

### 3.2 $l^1$ -norm Regularisation

Michailovich and Adam [1] proposed modelling the elements of  $\mathbf{x}$  as independent and identically distributed random variables obeying a Laplacian distribution instead of a Gaussian distribution on the grounds that the heavier tails of the Laplacian distribution permits better recovery of the stronger reflectors at structural boundaries. Their definition of the Laplacian distribution in [1] was for a real-valued random variable, but here we define an equivalent distribution for a complex-valued random variable.

Defining  $x_i = (\mathbf{x})_i$ , we define the real and imaginary parts of each  $x_i$  to be uncorrelated and to have a variance of  $\sigma_x^2$  each. The variance of each  $x_i$  is then  $2\sigma_x^2$  and we assign the following probability density function to each  $x_i$ ,

$$p(x_i) \propto \exp\left(-\frac{\sqrt{3}}{\sigma_x} |x_i|\right). \quad (15)$$

This probability density function is circularly symmetric in the complex plane, which keeps the phase of each  $x_i$  uniformly distributed between 0 and  $2\pi$ . Although the real and imaginary parts of each  $x_i$  are uncorrelated, they are not, strictly speaking, independent.

Assuming the components of  $\mathbf{x}$  to be independent and identically distributed, the probability density function in (15) gives rise to a regularising constraint on the  $l^1$ -norm instead of the  $l^2$ -norm, and our MAP estimation becomes

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left( \frac{1}{2\sigma_n^2} \|\mathbf{y} - H\mathbf{x}\|^2 + \frac{\sqrt{3}}{\sigma_x} \sum_{i=1}^N |x_i| \right). \quad (16)$$

## 4 Expectation-Maximisation (EM)

### 4.1 An Alternative Model for Tissue Reflectivity

So far, we have not addressed the problem of estimating the parameters (in particular, the covariance matrix of  $\mathbf{x}$ ) in the regularising constraints of (14) and (16). The reflectivities of biological tissues typically exhibit non-stationary statistics and so assuming a constant variance over the entire image is unlikely to yield optimal results. We previously stated that the echogenicities of soft tissues vary macroscopically in a piecewise smooth way but their reflectivities behave in a pseudo-random way on a microscopic level. This leads us to suggest that there is very little correlation between different samples of the tissue reflectivity and that the variance of each sample is determined by the macroscopic piecewise smooth echogenicity.

A simple way to express this behaviour is to model the tissue reflectivity as a field of uncorrelated and identically distributed random variables weighted by an echogenicity map (see Fig. 1). Let  $S$  be an  $N \times N$  diagonal matrix of real-valued non-negative samples of the echogenicity and let  $\mathbf{w}$  be an  $N \times 1$  vector of uncorrelated random variables of unit variance. We propose writing the tissue reflectivity as

$$\mathbf{x} = S \mathbf{w} \quad (17)$$

and we model  $\mathbf{w}$  as a complex-valued zero-mean Gaussian vector with covariance matrix  $E(\mathbf{w}\mathbf{w}^H) = 2I_N$  (this is consistent with the model in [15]). Since the echogenicity is piecewise-smooth and therefore likely to be well sparsified in the wavelet domain, we propose a wavelet-domain prior for the diagonal elements of  $S$ .

With this simple model for  $\mathbf{x}$ , we may write

$$p(\mathbf{x} | S) \propto \frac{1}{|S^2|} \exp\left(-\frac{1}{2} \mathbf{x}^H S^{-2} \mathbf{x}\right). \quad (18)$$

At this point, we make the following two important observations:

- If  $S$  were known exactly, then  $\mathbf{x}$  would just be a complex-valued Gaussian random vector with covariance matrix  $E(\mathbf{x}\mathbf{x}^H) = 2S^2$  and an estimate of  $\mathbf{x}$  could be calculated from  $\mathbf{y}$  by applying the Wiener filter as given in (14).
- Conversely, if  $\mathbf{x}$  were known exactly, then we could estimate  $S$  by treating  $\mathbf{w}$  as multiplicative noise and applying a suitable wavelet-domain denoising procedure to get rid of it.

These observations suggest that we can form an iterative image restoration algorithm by alternating between Wiener filtering to estimate  $\mathbf{x}$  and wavelet-domain denoising to estimate  $S$  to get successively better estimates of both. This approach forms the basic structure of our image restoration method, and to derive exact update rules for  $\mathbf{x}$  and  $S$ , we phrase our problem in terms of MAP estimation with hidden parameters and invoke the EM algorithm.

### 4.2 The EM Algorithm

The EM algorithm is an iterative procedure for maximum likelihood (ML) or MAP estimation in problems where there is an unobservable nuisance parameter in which we are not interested. It provides an alternative to explicit

marginalisation of the joint likelihood/posterior distribution which is often not analytically tractable. Each iteration of EM yields an estimate of the parameter of interest which increases its marginal likelihood/posterior and convergence to a local maximum of the marginal likelihood/posterior distribution is guaranteed.

We present, without proof, the mechanics of the EM algorithm for MAP estimation as described in [23]. Let  $\Theta$  be the parameter of interest, let  $\mathbf{U}$  be the observed data from which we wish to estimate  $\Theta$ , let  $\mathbf{J}$  be the nuisance parameter and let  $\hat{\Theta}_k$  be the estimate of  $\Theta$  at the  $k$ th iteration. At the  $k$ th iteration, we execute:

- The E-step: We calculate the expected joint log-likelihood of  $\mathbf{U}$  and  $\mathbf{J}$  given  $\Theta$ ,

$$Q\left(\Theta \mid \hat{\Theta}_k\right) = \mathbb{E}\left[\ln p(\mathbf{U}, \mathbf{J} \mid \Theta) \mid \mathbf{U}, \hat{\Theta}_k\right] = \int p(\mathbf{J} \mid \mathbf{U}, \hat{\Theta}_k) \ln p(\mathbf{U}, \mathbf{J} \mid \Theta) d\mathbf{J}. \quad (19)$$

- The M-step: We calculate the next estimate of  $\Theta$ ,

$$\hat{\Theta}_{k+1} = \arg \max_{\Theta} \left[ Q\left(\Theta \mid \hat{\Theta}_k\right) + \ln p(\Theta) \right]. \quad (20)$$

For our image restoration problem, we have  $\mathbf{U} = \mathbf{y}$ ; since we are able to specify a prior explicitly (in the wavelet domain) for  $S$ , and since the log-prior  $\ln p(\Theta)$  appears explicitly in the M-step, it makes sense to assign  $\Theta = S$ , which leaves  $\mathbf{J} = \mathbf{x}$ .

So far, we have implicitly assumed the variance  $2\sigma_n^2$  of the additive noise term in (7) to be known *a priori*. While robust estimators exist for determining the additive noise variance at the start of the algorithm, we can also update our estimate of the additive noise variance in each iteration of the algorithm by including  $\sigma_n$  as a parameter of interest to be estimated. We introduce a flat prior for  $\sigma_n$  which corresponds to solving for its ML estimate.

### 4.3 E-step: Wiener Filtering

We define  $\hat{S}_k$  and  $\hat{\sigma}_{n,k}$  to be the estimates of  $S$  and  $\sigma_n$  respectively at the  $k$ th iteration. Substituting  $\mathbf{U} = \mathbf{y}$ ,  $\mathbf{J} = \mathbf{x}$  and  $\Theta = \{S, \sigma_n\}$  into (19), we obtain

$$Q\left(S, \sigma_n \mid \hat{S}_k, \hat{\sigma}_{n,k}\right) = \int p(\mathbf{x} \mid \mathbf{y}, \hat{S}_k, \hat{\sigma}_{n,k}) \ln p(\mathbf{y}, \mathbf{x} \mid S, \sigma_n) d\mathbf{x}. \quad (21)$$

To obtain a closed-form expression for this integral, we first turn our attention to finding an expression for the ‘nuisance posterior’  $p(\mathbf{x} \mid \mathbf{y}, \hat{S}_k, \hat{\sigma}_{n,k})$ . Applying Bayes’s rule and recognising that  $p(\mathbf{y} \mid \mathbf{x}, \hat{S}_k, \hat{\sigma}_{n,k}) = p(\mathbf{y} \mid \mathbf{x}, \hat{\sigma}_{n,k})$  and  $p(\mathbf{x} \mid \hat{S}_k, \hat{\sigma}_{n,k}) = p(\mathbf{x} \mid \hat{S}_k)$ , we have

$$p(\mathbf{x} \mid \mathbf{y}, \hat{S}_k, \hat{\sigma}_{n,k}) = \frac{p(\mathbf{y} \mid \mathbf{x}, \hat{S}_k, \hat{\sigma}_{n,k}) p(\mathbf{x} \mid \hat{S}_k, \hat{\sigma}_{n,k})}{p(\mathbf{y} \mid \hat{S}_k, \hat{\sigma}_{n,k})} = \frac{p(\mathbf{y} \mid \mathbf{x}, \hat{\sigma}_{n,k}) p(\mathbf{x} \mid \hat{S}_k)}{p(\mathbf{y} \mid \hat{S}_k, \hat{\sigma}_{n,k})}. \quad (22)$$

The denominator in the rightmost term of (22) does not depend on  $\mathbf{x}$  and may be regarded as just a normalisation constant. Substituting (12) and (18) into the numerator and simplifying, we obtain

$$p(\mathbf{x} \mid \mathbf{y}, \hat{S}_k, \hat{\sigma}_{n,k}) \propto \exp \left\{ -\frac{1}{2\hat{\sigma}_{n,k}^2} \left[ \mathbf{x}^H \left( H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2} \right) \mathbf{x} - 2 \operatorname{Re}(\mathbf{x}^H H^H \mathbf{y}) \right] \right\} \quad (23)$$

which is a multivariate complex Gaussian distribution with covariance matrix  $C_k$  and mean  $\mathbf{m}_k$  given by

$$C_k = \text{Var} \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = 2\widehat{\sigma}_{n,k}^2 \left( H^H H + \widehat{\sigma}_{n,k}^2 \widehat{S}_k^{-2} \right)^{-1}, \quad (24)$$

$$\mathbf{m}_k = \text{E} \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = (2\widehat{\sigma}_{n,k}^2)^{-1} C_k H^H \mathbf{y} = \left( H^H H + \widehat{\sigma}_{n,k}^2 \widehat{S}_k^{-2} \right)^{-1} H^H \mathbf{y}. \quad (25)$$

The mean  $\mathbf{m}_k$  is the minimum mean squared error (MMSE) estimate of  $\mathbf{x}$  given  $\widehat{S}_k$  and therefore represents our best estimate of  $\mathbf{x}$  at the  $k$ th iteration. Comparing the rightmost expression in (25) with (14) shows that calculation of  $\mathbf{m}_k$  is just the same as Wiener filtering with  $C_x = \widehat{S}_k^2$ .

Returning now to the calculation of the integral in (21), we note that  $p(\mathbf{y}, \mathbf{x} \mid S, \sigma_n) = p(\mathbf{y} \mid \mathbf{x}, \sigma_n) p(\mathbf{x} \mid S) \Rightarrow \ln p(\mathbf{y}, \mathbf{x} \mid S, \sigma_n) = \ln p(\mathbf{y} \mid \mathbf{x}, \sigma_n) + \ln p(\mathbf{x} \mid S)$ , and we may write

$$Q \left( S, \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = Q_1 \left( S \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) + Q_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right), \quad (26)$$

$$Q_1 \left( S \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = \int p \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) \ln p \left( \mathbf{x} \mid S \right) d\mathbf{x}, \quad (27)$$

$$Q_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = \int p \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) \ln p \left( \mathbf{y} \mid \mathbf{x}, \sigma_n \right) d\mathbf{x}. \quad (28)$$

To expand  $Q_1 \left( S \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right)$ , we substitute (18) into (27) and discard constants to yield

$$\widetilde{Q}_1 \left( S \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = -2 \ln |S| - \frac{1}{2} \int \mathbf{x}^H S^{-2} \mathbf{x} p \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) d\mathbf{x}. \quad (29)$$

Recognising that the integral on the right-hand side is just the conditional expectation  $\text{E} \left( \mathbf{x}^H S^{-2} \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right)$ , we follow a similar calculation in [20, section 15.4] and simplify as follows,

$$\begin{aligned} \text{E} \left( \mathbf{x}^H S^{-2} \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) &= \text{E} \left[ \text{Tr} \left( S^{-2} \mathbf{x} \mathbf{x}^H \right) \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right] = \text{Tr} \left[ S^{-2} \text{E} \left( \mathbf{x} \mathbf{x}^H \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) \right] \\ &= \text{Tr} \left[ S^{-2} \left( C_k + \mathbf{m}_k \mathbf{m}_k^H \right) \right] = \text{Tr} \left( S^{-2} C_k \right) + \mathbf{m}_k^H S^{-2} \mathbf{m}_k, \end{aligned} \quad (30)$$

and substituting this expression back into (29), we obtain

$$\widetilde{Q}_1 \left( S \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = -2 \ln |S| - \frac{1}{2} \left[ \text{Tr} \left( S^{-2} C_k \right) + \|S^{-1} \mathbf{m}_k\|^2 \right]. \quad (31)$$

To expand  $Q_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right)$ , we substitute (12) into (28) and discard constants to yield

$$\widetilde{Q}_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = -2N \ln \sigma_n - \frac{1}{2\sigma_n^2} \int \|\mathbf{y} - H\mathbf{x}\|^2 p \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) d\mathbf{x}. \quad (32)$$

As before, we recognise that the integral on the right-hand side is just the conditional expectation

$$\text{E} \left( \|\mathbf{y} - H\mathbf{x}\|^2 \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) = \mathbf{y}^H \mathbf{y} - 2\text{Re} \left[ \mathbf{y}^H H \text{E} \left( \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) \right] + \text{E} \left( \mathbf{x}^H H^H H \mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \widehat{\sigma}_{n,k} \right) \quad (33)$$

and, evaluating each of these expectations, we obtain

$$\begin{aligned} \widetilde{Q}_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) &= -2N \ln \sigma_n - \frac{1}{2\sigma_n^2} \left[ \mathbf{y}^H \mathbf{y} - 2\text{Re} \left( \mathbf{y}^H H \mathbf{m}_k \right) + \mathbf{m}_k^H H^H H \mathbf{m}_k + \text{Tr} \left( H^H H C_k \right) \right] \\ &= -2N \ln \sigma_n - \frac{1}{2\sigma_n^2} \left[ \|\mathbf{y} - H\mathbf{m}_k\|^2 + \text{Tr} \left( H^H H C_k \right) \right]. \end{aligned} \quad (34)$$

We conclude our discussion on the E-step with the important observation that, because our joint log-likelihood  $Q \left( S, \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right)$  is a sum of a function purely of  $S$  and a function purely of  $\sigma_n$ , and because  $S$  is independent

of  $\sigma_n$ , we can split our M-step into two separate optimisation problems,

$$\begin{aligned}\widehat{S}_{k+1} &= \arg \max_S \left[ \widetilde{Q}_1 \left( S \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) + \ln p(S) \right] \\ &= \arg \max_S \left\{ -2 \ln |S| - \frac{1}{2} \left[ \text{Tr} (S^{-2} C_k) + \|S^{-1} \mathbf{m}_k\|^2 \right] + \ln p(S) \right\},\end{aligned}\quad (35)$$

$$\begin{aligned}\widehat{\sigma}_{n,k+1} &= \arg \max_{\sigma_n} \widetilde{Q}_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right) \\ &= \arg \max_{\sigma_n} \left\{ -2N \ln \sigma_n - \frac{1}{2\sigma_n^2} \left[ \|\mathbf{y} - H \mathbf{m}_k\|^2 + \text{Tr} (H^H H C_k) \right] \right\},\end{aligned}\quad (36)$$

and we discuss each optimisation problem individually in each of the next two sections.

#### 4.4 M-step: Logarithmic Denoising

Before we launch into the derivation of the update rule for  $\widehat{S}$  in the M-step, we detour briefly to discuss the problem of estimating  $S$  given  $\mathbf{x}$  by treating  $\mathbf{w}$  in (17) as multiplicative noise to be removed. We define  $x_i = (\mathbf{x})_i$ ,  $s_i = (S)_{ii}$  and  $w_i = (\mathbf{w})_i$  and we rewrite (17) component-wise as  $x_i = s_i w_i$  (recall that  $x_i$  and  $w_i$  are complex-valued whereas  $s_i$  is real-valued and non-negative). To turn  $\{w_i\}$  into additive noise, we take the logarithms of the moduli of both sides; defining  $\widetilde{x}_i = \ln |x_i|$ ,  $\widetilde{s}_i = \ln s_i$  and  $\widetilde{w}_i = \ln |w_i|$ , we obtain

$$\widetilde{x}_i = \widetilde{s}_i + \widetilde{w}_i, \quad i = 1, \dots, N. \quad (37)$$

The logarithmic noise term  $\{\widetilde{w}_i\}$  has probability density function, mean and variance given by

$$p(\widetilde{w}_i) = \exp \left[ 2\widetilde{w}_i - \frac{1}{2} \exp(2\widetilde{w}_i) \right], \quad \text{E}(\widetilde{w}_i) = \frac{1}{2} (\ln 2 - \gamma) \approx 0.0580, \quad \text{Var}(\widetilde{w}_i) = \frac{\pi^2}{24} \approx 0.4112, \quad (38)$$

where  $\gamma$  is the Euler-Mascheroni constant and has an approximate value of 0.5772. The derivation of this probability density function and the calculation of its mean and variance are detailed in Appendix A.

Since the echogenicity is piecewise smooth, we expect the log-echogenicity  $\{\widetilde{s}_i\}$  to also be piecewise smooth and to have a sparse representation in the wavelet domain. We can therefore denoise by applying wavelet shrinkage: since most of the energy of  $\{\widetilde{s}_i\}$  will be concentrated into just a few wavelet coefficients, we can modify the wavelet coefficients of  $\{\widetilde{x}_i\}$  according to some shrinkage rule that attenuates the logarithmic noise term  $\{\widetilde{w}_i\}$ . Most wavelet shrinkage rules are based on the assumption that the wavelet coefficients of the additive noise are Gaussian, which at first sight seems to be violated by the non-Gaussianity of the probability density function in (38). In practice, however, we have found that, because of the bandlimiting of each wavelet subband, the central limit theorem keeps the wavelet coefficients of  $\{\widetilde{w}_i\}$  approximately Gaussian. The additive noise is also usually assumed to have zero mean, so we need to subtract  $\frac{1}{2} (\ln 2 - \gamma)$  from  $\{\widetilde{x}_i\}$  before applying wavelet shrinkage.

It is well known that wavelet shrinkage corresponds to MAP estimation with a wavelet-domain prior, the exact form of which depends on the specific shrinkage rule used [24, 25]. Hence, we may regard the logarithmic denoising of  $\{\widetilde{x}_i\}$  to recover  $\{\widetilde{s}_i\}$  and estimate  $\{s_i\}$  as being equivalent to the MAP estimation problem

$$\begin{aligned}\widehat{S} &= \arg \max_S p(S \mid \mathbf{x}) = \arg \max_S p(\mathbf{x} \mid S) p(S) = \arg \max_S [\ln p(\mathbf{x} \mid S) + \ln p(S)] \\ &= \arg \max_S \left[ -2 \ln |S| - \frac{1}{2} \mathbf{x}^H S^{-2} \mathbf{x} + \ln p(S) \right] = \arg \max_S \left[ - \sum_{i=1}^N \left( 2 \ln s_i + \frac{|x_i|^2}{2s_i^2} \right) + \ln p(S) \right]\end{aligned}\quad (39)$$

when the prior  $p(S)$  is defined in terms of the wavelet coefficients of the log-echogenicity  $\{\widetilde{s}_i\}$ .

Returning now to the derivation of the update rule for  $\widehat{S}$ , if we define  $\sigma_{i,k}^2 = [C_k]_{ii}$  and  $m_{i,k} = [\mathbf{m}_k]_i$ , we may

rewrite (35) as

$$\widehat{S}_{k+1} = \arg \max_S \left[ - \sum_{i=1}^N \left( 2 \ln s_i + \frac{|m_{i,k}|^2 + \sigma_{i,k}^2}{2s_i^2} \right) + \ln p(S) \right]. \quad (40)$$

If we now compare this form of the update rule with (39), we see that the two expressions are identical simply by letting  $x_i^2 = |m_{i,k}|^2 + \sigma_{i,k}^2$ , and in light of our previous discussion on logarithmic denoising, we conclude that we can calculate our next estimate of  $S$  simply by applying wavelet shrinkage to  $\left\{ \ln \left( \sqrt{|m_{i,k}|^2 + \sigma_{i,k}^2} \right) \right\}$ .

As an initial estimate, we suggest setting all the diagonal elements of  $S$  to an estimate of the global blurred noise-to-signal ratio of the image, calculated as the ratio of an estimate of the additive noise variance to  $N^{-1} \|\mathbf{y}\|^2$ . This makes the E-step in the first iteration of the algorithm identical to zero-order Tikhonov regularisation.

## 4.5 M-Step: Updating the Additive Noise Variance Estimate

Differentiating the log-likelihood  $\widetilde{Q}_2 \left( \sigma_n \mid \widehat{S}_k, \widehat{\sigma}_{n,k} \right)$  in (34) yields

$$\frac{d\widetilde{Q}_2}{d\sigma_n} = -\frac{2N}{\sigma_n} + \frac{1}{\sigma_n^3} \left[ \text{Tr} \left( H^H H C_k \right) + \|\mathbf{y} - H\mathbf{m}_k\|^2 \right], \quad (41)$$

and setting the derivative to zero, we obtain

$$\widehat{\sigma}_{n,k+1}^2 = \frac{1}{2N} \left[ \text{Tr} \left( H^H H C_k \right) + \|\mathbf{y} - H\mathbf{m}_k\|^2 \right]. \quad (42)$$

The second derivative of the cost function at this value of  $\sigma_n$  is given by

$$\left. \frac{d^2\widetilde{Q}_2}{d\sigma_n^2} \right|_{\sigma_n = \widehat{\sigma}_{n,k+1}} = -8N^2 \left[ \text{Tr} \left( H^H H C_k \right) + \|\mathbf{y} - H\mathbf{m}_k\|^2 \right]^{-1} < 0 \quad (43)$$

which confirms that a local maximum has been reached.

At the start of the algorithm, we suggest taking the discrete wavelet transform (DWT) of the noisy image  $\mathbf{y}$  and initialising our estimate of  $\sigma_n$  to the median of the moduli of the finest-scale coefficients divided by  $\sqrt{2 \ln 2}$ . This estimator for  $\sigma_n$  is based on the assumption that the finest-scale coefficients are dominated by the additive noise  $\mathbf{n}$  and is robust to the presence of outliers. We explain the rationale of this estimator in greater detail in Appendix B.

## 4.6 Summary of the Algorithm

We summarise our EM algorithm for image restoration as follows (we include references to equations earlier in the paper and, for notational convenience, we drop the subscript  $k$ ):

Estimate  $\widehat{\sigma}_n^2$  with the robust wavelet estimator.

Initialise  $\widehat{S} = 2N\widehat{\sigma}_n^2 \|\mathbf{y}\|^{-2} I_N$ .

While termination condition is not satisfied

E-step:

Define  $C = 2\widehat{\sigma}_n^2 \left( H^H H + \widehat{\sigma}_n^2 \widehat{S}^{-2} \right)^{-1}$  (c.f. (24)).

Calculate  $\mathbf{m} = \left( 2\widehat{\sigma}_n^2 \right)^{-1} C H^H \mathbf{y}$  (c.f. (25)).

Calculate  $\sigma_i^2 = (C)_{ii}$  for  $i = 1, \dots, N$ .

M-step:

Calculate  $\widetilde{x}_i = \ln \left( \sqrt{|m_i|^2 + \sigma_i^2} \right) = \frac{1}{2} \ln \left( |m_i|^2 + \sigma_i^2 \right)$  for  $i = 1, \dots, N$ .

Estimate  $\{\tilde{s}_i\}$  by applying wavelet shrinkage to  $\{\tilde{x}_i + \frac{1}{2}(\gamma - \ln 2)\}$ .  
Calculate  $\hat{s}_i = \exp(\tilde{s}_i)$  for  $i = 1, \dots, N$ .  
Calculate  $\hat{\sigma}_n^2 = (2N)^{-1} \left[ \|\mathbf{y} - H\mathbf{m}\|^2 + \text{Tr}(H^H H C) \right]$  (c.f. (42)).

end

Once the algorithm has terminated, we take  $\mathbf{m}$  as our Bayesian estimate of the tissue reflectivity and  $\{\hat{s}_i\}$  as our MAP estimate of the echogenicity.

## 5 Computational Considerations

### 5.1 The Conditional Covariance Matrix $C_k$

In Section 2.2, we stated that the size of the blurring matrix  $H$  is typically too large to be explicitly computed and we described methods to approximate multiplications by  $H$  and  $H^H$  based on locally shift-invariant blurring operators. It follows that the direct matrix inversion required to calculate the conditional covariance matrix  $C_k$  is also practically infeasible, but we note that our algorithm does not require explicit calculation of  $C_k$ ; it only requires multiplication of a vector by  $C_k$  and extraction of the diagonal elements of  $C_k$ .

The first occurrence of  $C_k$  in our algorithm is in the expression for  $\mathbf{m}_k$  in (25). We recognise that calculating  $\mathbf{m}_k$  is the same as solving the symmetric positive definite system  $(H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2}) \mathbf{m}_k = H^H \mathbf{y}$ , which we can do iteratively using a gradient-based method such as the conjugate gradients algorithm [26, section 2.3.1],[27, section 2.7]; this only requires multiplication by  $(H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2})$  which is feasible. In our experiments, we used the conjugate gradients algorithm with diagonal preconditioning, i.e. at each iteration of the algorithm, we preconditioned the residual by dividing each of its components by the corresponding diagonal element in  $(H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2})$ . The diagonal elements of  $\hat{\sigma}_{n,k}^2 \hat{S}_k^{-2}$  are trivial to calculate (since  $\hat{S}_k$  is diagonal), and we approximate each diagonal element of  $H^H H$  by the energy of the impulse response of the region within which the corresponding image sample lies.

To decide when to terminate the conjugate gradients algorithm, we used the following termination rule adopted from [26, section 4.2.1],

$$\left\| (H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2}) \mathbf{m}_k - H^H \mathbf{y} \right\| \leq \tau \left( \left\| H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2} \right\|_F \|\mathbf{m}_k\| + \|H^H \mathbf{y}\| \right) \quad (44)$$

where  $\tau > 0$  is a stopping tolerance. We make the following important remarks:

- The Frobenius norm  $\left\| H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2} \right\|_F$  can be expressed as  $\sqrt{\text{Tr} \left[ (H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2})^2 \right]}$ . Calculating it exactly is difficult, so instead we used the following unbiased stochastic estimator taken from [28, 29]: we generated a random  $N \times 1$  vector  $\mathbf{u}$  whose elements may be either 1 or  $-1$  with equal probability, and we made the approximation

$$\text{Tr} \left[ (H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2})^2 \right] = \text{E} \left\{ \mathbf{u}^T (H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2})^2 \mathbf{u} \right\} \approx \mathbf{u}^T (H^H H + \hat{\sigma}_{n,k}^2 \hat{S}_k^{-2})^2 \mathbf{u}. \quad (45)$$

- Selection of an appropriate stopping tolerance  $\tau$  turns out to be quite crucial. A value of  $\tau$  that is too small increases the computation time of each E-step and can, as we discovered from *in vitro* experiments, exacerbate the effect of errors in our approximation of the blurring operator. Conversely, selecting a value of  $\tau$  that is too large results in under-filtered images being passed to the logarithmic denoising stage and introduces artifacts into the end result. In our experiments, we chose appropriate values of  $\tau$  by trial-and-error.

The second occurrence of  $C_k$  is in (40), where its diagonal elements  $\{\sigma_{i,k}^2\}$  are needed to update the echogenicity estimate. Since we are unable, in practice, to explicitly specify  $C_k$ , we cannot access its diagonal elements  $\{\sigma_{i,k}^2\}$  directly. We recognise, however, that each  $\sigma_{i,k}^2$  is just the variance of  $(\mathbf{x})_i$  given  $\mathbf{y}$ ,  $\widehat{S}_k$  and  $\widehat{\sigma}_{n,k}$ . We can assume local ergodicity and estimate  $\{\sigma_{i,k}^2\}$  by computing sample variances over local neighbourhoods in some estimate of  $\mathbf{x}$ . Since  $\mathbf{m}_k$  is our best estimate of  $\mathbf{x}$  given  $\mathbf{y}$ ,  $\widehat{S}_k$  and  $\widehat{\sigma}_{n,k}$ , we expect a good estimate of  $\{\sigma_{i,k}^2\}$  to be obtained by multiplying each component of  $\mathbf{m}_k$  by its complex conjugate and convolving the resulting image with a suitably-scaled rectangular kernel.

The third occurrence of  $C_k$  is in the update rule (42) for  $\widehat{\sigma}_{n,k}$ , in the term  $\text{Tr}(H^H H C_k)$ . We can use an approximation similar to (45) and write

$$\text{Tr}(H^H H C_k) = \text{E}[\mathbf{u}^T H^H H C_k \mathbf{u}] \approx \mathbf{u}^T H^H H C_k \mathbf{u} = 2\widehat{\sigma}_{n,k}^2 \mathbf{u}^T H^H H (H^H H + \widehat{\sigma}_{n,k}^2 \widehat{S}_k^{-2})^{-1} \mathbf{u}. \quad (46)$$

Multiplication by  $(H^H H + \widehat{\sigma}_{n,k}^2 \widehat{S}_k^{-2})^{-1}$  has to be done via the conjugate gradients algorithm, and so this step can be quite time-consuming.

## 5.2 Choice of Wavelet Transform and Shrinkage Rule

In our treatment of the logarithmic denoising portion of the M-step, we did not specify the form of the log prior  $\ln p(S)$  in (39) and (40) beyond stating that it is to be specified in terms of the wavelet coefficients of the echogenicity image. We are therefore free to use any wavelet transform and any shrinkage rule we choose.

We recommend using the dual-tree complex wavelet transform (DTCWT) [30, 31, 32] over the conventional DWT because it has better properties than the DWT for image processing which come with only a small computational penalty and modest redundancy. For a real-valued signal, the conventional DWT uses a single dyadic filter tree to generate real-valued coefficients whereas the DTCWT uses a pair of dyadic filter trees to generate the real and imaginary parts of complex-valued coefficients. For an  $n$ -dimensional real-valued signal, the redundancy of the DTCWT is  $2^n : 1$ , i.e. the number of complex-valued coefficients is  $2^{n-1}$  times the number of samples in the signal. An extension of the DTCWT to complex-valued signals is discussed in [30].

Unlike the DWT, the DTCWT is redundant and hence not orthonormal; it is, however, energy-preserving and forms an almost-tight frame, i.e. the total energy of the DTCWT coefficients is virtually the same as the energy of the input signal. The DTCWT has the following advantages over the DWT:

- *Shift-invariance*: The magnitudes of the DTCWT coefficients remain approximately constant despite spatial shifts in the input signal, unlike the magnitudes of the DWT coefficients which can fluctuate significantly with spatial shifts.
- *Directional selectivity*: In two dimensions or higher, the DTCWT is able to localise features of opposing orientations into different subbands, which is not possible with the DWT because the coefficients of the DWT are real-valued.

For the logarithmic denoising, we propose using the bivariate shrinkage rule developed by Sendur and Selesnick [24, 25] which, when used in conjunction with the DTCWT, gives state-of-the-art denoising performance. This shrinkage rule attenuates each noisy complex-valued wavelet coefficient based on its magnitude and the magnitude of its parent (i.e. the noisy coefficient at the same spatial location in the same subband at the next coarser level).

## 6 Experimental Results

### 6.1 Simulation Results

We generated artificial tissue reflectivities from natural images of biological structures by multiplying them with complex-valued white Gaussian noise. The logarithm of each image was histogram equalised and scaled to give a signal-to-noise ratio in the logarithmic domain of 20dB. A set of eight impulse responses (for eight equally partitioned axial regions in image space) were generated by simulating the response of a linear array 6.5MHz probe to appropriately-positioned point scatterers in Field II [33]; each impulse response was then demodulated according to (4). The impulse responses were scaled to give the impulse response closest to the focus (which was approximately in the centre of the image) unit energy. We blurred the tissue reflectivities in the manner described in Section 3 and added complex-valued white Gaussian noise to give the resulting images a blurred signal-to-noise ratio (BSNR) of 20dB.

We tested our proposed EM algorithm and compared its performance to zero-order Tikhonov regularisation,  $l^1$ -norm (Laplacian) regularisation and ForWaRD. To perform the necessary function minimisations, we used the conjugate gradients algorithm with diagonal preconditioning; in the case of  $l^1$ -norm regularisation, this meant dividing elements of the gradient vector by the diagonals of the Hessian matrix, which vary from point to point (for a discussion on gradient vectors and matrices of real-valued functions of complex variables, see Appendix C). We quantified the performance of each algorithm according to its result's improvement in signal-to-noise ratio (ISNR) in dB, calculated as

$$\text{ISNR} = 20 \log_{10} \left( \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{y} - \mathbf{x}\|} \right) \quad (47)$$

where  $\mathbf{x}$  is the vectorised tissue reflectivity,  $\mathbf{y}$  is the noisy and blurred image and  $\hat{\mathbf{x}}$  is the vectorised restored image. We also tested our EM algorithm omitting the update rule for  $\hat{\sigma}_{n,k}$  to examine the difference in performance. All algorithms were run in Matlab 7 on a personal computer with a 3.2GHz processor and 1GB of memory.

#### Heart Image

Our first set of simulations was conducted on an image of a cross-section of a human heart. The results of the various restoration methods are shown in Fig. 2(a), and a plot of the evolution of ISNR against execution time for the zero-order Tikhonov, Laplacian and EM restorations is shown in Fig. 2(b) (we have omitted ForWaRD from this plot because its ISNR can only be sensibly computed at the end, after wavelet shrinkage). The EM algorithm, with and without updating  $\hat{\sigma}_{n,k}$ , is seen to outperform every other method by at least 0.63dB in ISNR and, qualitatively, it is clear that the EM solutions enjoy significantly better contrast than the other restored images. Inclusion of the update rule for  $\hat{\sigma}_{n,k}$  appears to have made very little difference to the ISNR of the EM solution, although it has significantly increased the computation time of each E-step, as can be seen from Fig. 2(b).

#### Kidney Image

Our second set of simulations was conducted on an image of a cross-section of a kidney. The results of the various restoration methods are displayed in Fig. 3(a), and we have a similar plot of the evolution of ISNR against execution time in Fig. 3(b). These results are consistent with our previous results, with the EM algorithm outperforming every other method by at least 0.61dB and its images exhibiting better contrast than the other solutions. It is encouraging to see that the echogenicity estimates from the EM algorithm have good visual quality: they are free of speckle and have good contrast and well-defined edges.

## 6.2 In Vitro Results

We acquired an image of a phantom containing spherical inclusions of different echogenicities embedded in a background of dense scatterers with a linear array 5-10MHz probe and a Dynamic Imaging Diasus ultrasound machine. The probe’s aperture contained 127 elements (with an active aperture of 32 elements) and spanned a lateral length of 40mm. The probe was operated with a single lateral focus at an axial depth of 26mm on both transmission and reception. Elevational focussing was achieved with an acoustic lens that had a focal length of 23mm. Standard delay-and-sum beamforming was applied and the beamformed traces were sampled at a rate of 66.6MHz. Each sampled trace was then demodulated to baseband, low-pass filtered and downsampled by a factor of nine. We calculated 16 impulse responses (one for each of 16 equally partitioned axial regions) by simulating the response to appropriately-positioned point scatterers in Field II. Each impulse response was then demodulated according to (4) and scaled so that the impulse response closest to the lateral focus had unit energy. We approximated the global shift-variant blurring operator by linearly interpolating between the simulated impulse responses.

As with the simulations, we tested our proposed EM algorithm (with and without updating  $\hat{\sigma}_{n,k}$ ) and compared its performance to zero-order Tikhonov regularisation, ForWaRD and Laplacian regularisation. Our approximation of the shift-variant blurring operator with locally shift-invariant impulse responses gave rise to artifacts in the restored images which we corrected using a method described in Appendix D. The restored images after artifact correction are shown in Fig. 4. The first thing to note is that the effects of restoration are most pronounced at the axial extremes of the image where lateral focussing is worst: the shapes of the spheres at the top and at the bottom of the image have been successfully corrected and appear circular in all of the restored images. To quantify the relative merits of the various restoration schemes, we calculated the contrasts between different pairs of adjacent regions in each image; these contrasts (in dB) are listed in Table 3. We note that, for a given pair of adjacent regions, the contrasts of the EM solutions are at least comparable to and often significantly stronger than the contrasts of the other restored images. Once again, we find it encouraging that the echogenicity estimates from the EM algorithm are free of speckle and have good contrast and well-defined boundaries.

## 7 Conclusions

In this paper, we have addressed the problem of image restoration in the context of pulse-echo ultrasound with the implicit assumption that the blurring operator is known *a priori*. The performance of an image restoration algorithm is highly dependent on the appropriateness of the underlying image model it assumes. The reflectivities of soft tissues imaged by pulse-echo ultrasound have substantially different statistics to the piecewise smooth natural images considered in mainstream image processing and models for such natural images cannot be applied directly to pulse-echo ultrasound. We observed that, though the echogenicities of soft tissues vary macroscopically in a piecewise smooth way, the reflectivities exhibit a pseudo-random texture on a microscopic scale due to the presence of small, weakly-scattering particles. We therefore proposed modelling reflectivity as the product of a piecewise smooth echogenicity map and a field of uncorrelated and identically distributed random variables.

The explicit separation of the reflectivity into an echogenicity component and a random component allows the echogenicity to be modelled as a natural image which can be sparsely represented in the wavelet domain. By assigning a Gaussian distribution to the random component and applying the EM algorithm, we were able to derive a restoration algorithm that consists of alternating between Wiener filtering to estimate the reflectivity and wavelet-based denoising to estimate the echogenicity. An alternative way of viewing this algorithm is to consider it as a Wiener filter with the signal variances updated occasionally by wavelet-based denoising. Because this restoration algorithm also yields an estimate of the echogenicity (which is free of speckle), it effectively solves both the image restoration problem and the speckle removal problem simultaneously.

In the derivation of our image restoration algorithm, we also incorporated a rule in the M-step for updating the variance of the additive white Gaussian noise term that accounts for measurement error. We tested our restoration algorithm with and without this extra update rule on images generated in simulation and acquired *in vitro* and we compared the results from our algorithm with restorations based on  $l^1$ - and  $l^2$ -norm regularisation and ForWaRD. The simulation results indicated that our algorithm is capable of producing restored images with greater fidelity to the true tissue reflectivity (as measured by ISNR) and visually we observed our algorithm's results to have better local contrast. The improvement in local contrast can also be seen in the *in vitro* results, where the restored images from our algorithm exhibited comparable and often significantly better local contrast than the other restored images. Inclusion of the update rule for the variance of the additive noise appears to have made little difference to image quality but significantly increases computation time; therefore, we recommend omitting this update rule unless the initial estimate of this variance is known to be unreliable.

We conclude that our image restoration algorithm for pulse-echo ultrasound is competitive with the current state-of-the-art and can produce results that have superior image quality in terms of local contrast and fidelity to the true tissue reflectivity. When the rule for updating the additive noise variance is omitted, the M-step consists only of wavelet shrinkage (which is computationally very efficient) and the computational cost of our algorithm as a whole is dominated by the E-step which involves the same gradient-based optimisations as  $l^1$ - and  $l^2$ -norm regularisation. Hence, the benefits introduced by our algorithm come with only a small computational penalty.

We have also included in this paper closed-form expressions for the statistics of the logarithm of a Rayleigh-distributed random variable ((38) and Appendix A) and a robust wavelet-based estimator for the variance of complex-valued additive white Gaussian noise in a realistic signal (Section 4.5 and Appendix B). We have been unable to find these expressions in the technical literature and so we have presented them here for the sake of completeness.

## A Derivation of the Statistics of Logarithmic Noise

In this appendix we derive the probability density function and the statistics quoted in (38) for the logarithmic noise term  $\{\tilde{w}_i\}$  in (37). We first state the following two useful equations,

$$\int_0^{\infty} e^{-u} \ln u \, du = -\gamma, \quad (48)$$

$$\int_0^{\infty} e^{-u} (\ln u)^2 \, du = \gamma^2 + \frac{\pi^2}{6}, \quad (49)$$

where  $\gamma$  is the Euler-Mascheroni constant. (48) and (49) are, respectively, special cases of (4.331-1) and (4.335-1) in [34].

*Proposition:* Given a complex-valued Gaussian random variable  $Z$  with mean  $E(Z) = 0$  and variance  $E(|Z|^2) = 2$ , and defining  $\tilde{W} = \ln |Z|$ , the probability density function of  $\tilde{W}$  is given by  $p_{\tilde{W}}(\tilde{w}) = \exp[2\tilde{w} - \frac{1}{2} \exp(2\tilde{w})]$ .

*Proof:* Define  $W = |Z|$ . Marginalising the probability density function of  $Z$  over  $\arg(Z)$  (see, for example, [18]), it can be shown that  $W$  is distributed according to the Rayleigh distribution,

$$p_W(w) = \begin{cases} 0 & \text{if } w < 0, \\ w \exp\left(-\frac{w^2}{2}\right) & \text{if } w \geq 0. \end{cases}$$

Now define  $\tilde{W} = \ln W$ . The probability density function of  $\tilde{W}$  is therefore

$$p_{\tilde{W}}(\tilde{w}) = p_W(w) \left. \frac{dw}{d\tilde{w}} \right|_{w=e^{\tilde{w}}} = \exp(2\tilde{w}) \exp\left(-\frac{e^{2\tilde{w}}}{2}\right) = \exp\left[2\tilde{w} - \frac{1}{2} \exp(2\tilde{w})\right].$$

*Proposition:* Given a complex-valued Gaussian random variable  $Z$  with mean  $\mathbb{E}(Z) = 0$  and variance  $\mathbb{E}(|Z|^2) = 2$ , and defining  $\widetilde{W} = \ln |Z|$ , the mean of  $\widetilde{W}$  is given by  $\mathbb{E}(\widetilde{W}) = \frac{1}{2}(\ln 2 - \gamma)$ .

*Proof:* The mean of  $\widetilde{W}$  is just the expected value of  $\ln W$ ,

$$\begin{aligned}\mathbb{E}(\widetilde{W}) &= \int_0^\infty (\ln w) w \exp\left(-\frac{w^2}{2}\right) dw = \int_0^\infty (\ln \sqrt{2u}) e^{-u} du \\ &= \frac{1}{2} \int_0^\infty (\ln 2 + \ln u) e^{-u} du = \frac{\ln 2}{2} \int_0^\infty e^{-u} du + \frac{1}{2} \int_0^\infty e^{-u} \ln u du.\end{aligned}$$

The first integral evaluates to 1 and, from (48), the second integral evaluates to  $-\gamma$ . Hence,  $\mathbb{E}(\widetilde{W}) = \frac{1}{2}(\ln 2 - \gamma)$ .

*Proposition:* Given a complex-valued Gaussian random variable  $Z$  with mean  $\mathbb{E}(Z) = 0$  and variance  $\mathbb{E}(|Z|^2) = 2$ , and defining  $\widetilde{W} = \ln |Z|$ , the variance of  $\widetilde{W}$  is given by  $\text{Var}(\widetilde{W}) = \frac{\pi^2}{24}$ .

*Proof:* The second moment of  $\widetilde{W}$  is given by

$$\begin{aligned}\mathbb{E}(\widetilde{W}^2) &= \int_0^\infty (\ln w)^2 w \exp\left(-\frac{w^2}{2}\right) dw = \int_0^\infty (\ln \sqrt{2u})^2 e^{-u} du = \frac{1}{4} \int_0^\infty (\ln 2 + \ln u)^2 e^{-u} du \\ &= \frac{1}{4} \left[ (\ln 2)^2 \int_0^\infty e^{-u} du + 2 \ln 2 \int_0^\infty e^{-u} \ln u du + \int_0^\infty e^{-u} (\ln u)^2 du \right].\end{aligned}$$

As before, the first integral evaluates to 1, and, from (48) and (49), the second and third integrals evaluate to  $-\gamma$  and  $\gamma^2 + \frac{\pi^2}{6}$  respectively. So  $\mathbb{E}(\widetilde{W}^2) = \frac{1}{4} \left[ (\ln 2)^2 - 2\gamma \ln 2 + \gamma^2 + \frac{\pi^2}{6} \right]$ , and

$$\text{Var}(\widetilde{W}) = \mathbb{E}(\widetilde{W}^2) - \left[ \mathbb{E}(\widetilde{W}) \right]^2 = \frac{1}{4} \left[ (\ln 2)^2 - 2\gamma \ln 2 + \gamma^2 + \frac{\pi^2}{6} \right] - \frac{1}{4} \left[ (\ln 2)^2 - 2\gamma \ln 2 + \gamma^2 \right] = \frac{\pi^2}{24}.$$

## B Derivation of the Robust Noise Variance Estimator

In this appendix we derive the robust estimator for the variance of complex-valued additive white Gaussian noise (AWGN) introduced in Section 4.5. Our derivation is analogous to the derivation of a similar robust estimator for the variance of real-valued AWGN presented in [8, section 10.2.4].

*Proposition:* Given a complex-valued Gaussian random variable  $Z$  with mean  $\mathbb{E}(Z) = 0$  and variance  $\mathbb{E}(|Z|^2) = 2\sigma_n^2$ , and defining  $W = |Z|$ , the median of  $W$  is  $\sigma_n \sqrt{2 \ln 2}$ .

*Proof:* The calculation is fairly straightforward. As in the previous appendix, if we marginalise the probability density function of  $Z$  over  $\arg(Z)$ , we find that  $W$  is distributed according to the Rayleigh distribution

$$p_W(w) = \begin{cases} 0 & \text{if } w < 0, \\ \frac{w}{\sigma_n^2} \exp\left(-\frac{w^2}{2\sigma_n^2}\right) & \text{if } w \geq 0. \end{cases}$$

The median  $w_m$  of  $W$  satisfies

$$\int_{-\infty}^{w_m} p_W(w) dw = \frac{1}{2} \implies \int_0^{w_m} \frac{w}{\sigma_n^2} \exp\left(-\frac{w^2}{2\sigma_n^2}\right) dw = \frac{1}{2} \implies w_m = \sigma_n \sqrt{2 \ln 2}.$$

Given a realistic (i.e. bandlimited) signal contaminated by complex-valued additive white Gaussian noise, we expect its finest-scale DWT coefficients to be dominated by the noise, and since the DWT is orthonormal, the finest-scale coefficients are likely to also be distributed as complex-valued Gaussian random variables with the same variance as the noise. The sample median of the moduli of these coefficients is generally immune to any

potential outliers and is approximately equal to the population median  $\sigma_n \sqrt{2 \ln 2}$ , so a robust estimate of  $\sigma_n$  is this sample median divided by  $\sqrt{2 \ln 2}$ .

## C Gradient and Hessian of a Real-Valued Function of a Complex-Valued Vector

The form of the conjugate gradients algorithm which we used to perform function minimisation requires knowledge of the gradient vector and the Hessian matrix of the cost function to be minimised. The cost functions in (14) and (16) are real-valued functions of complex-valued vectors, and in this appendix we explain briefly how their gradient vectors and Hessian matrices were calculated. More detailed discussions of gradient vectors for real-valued functions of complex-valued vectors can be found in [35, Appendix B], [20, chapter 15] and [36].

A real-valued function of a complex-valued vector may be regarded as a function of the vector and its complex conjugate. Adopting the definition in [35, Appendix B], we define the partial derivative operators  $\frac{\partial}{\partial \mathbf{x}}$  and  $\frac{\partial}{\partial \mathbf{x}^*}$  with respect to a complex-valued vector  $\mathbf{x} \in \mathbb{C}^N$  to be

$$\frac{\partial}{\partial \mathbf{x}} \equiv \left[ \frac{\partial}{\partial x_1} \quad \cdots \quad \frac{\partial}{\partial x_N} \right]^T = \frac{1}{2} \left[ \frac{\partial}{\partial \text{Re}(x_1)} - j \frac{\partial}{\partial \text{Im}(x_1)} \quad \cdots \quad \frac{\partial f}{\partial \text{Re}(x_N)} - j \frac{\partial}{\partial \text{Im}(x_N)} \right]^T \quad (50)$$

$$\frac{\partial}{\partial \mathbf{x}^*} \equiv \left[ \frac{\partial}{\partial x_1^*} \quad \cdots \quad \frac{\partial}{\partial x_N^*} \right]^T = \frac{1}{2} \left[ \frac{\partial}{\partial \text{Re}(x_1)} + j \frac{\partial}{\partial \text{Im}(x_1)} \quad \cdots \quad \frac{\partial f}{\partial \text{Re}(x_N)} + j \frac{\partial}{\partial \text{Im}(x_N)} \right]^T \quad (51)$$

where  $x_i = (\mathbf{x})_i$  and  $\mathbf{x}^*$  is the complex conjugate of  $\mathbf{x}$ . Although, strictly speaking,  $\mathbf{x}$  and  $\mathbf{x}^*$  are not independent of each other, for the purposes of these partial derivative operators, we may treat them as independent variables.

The gradient  $\nabla \phi$  of a function  $\phi : \mathbf{x} \in \mathbb{C}^N \rightarrow \mathbb{R}$  is then defined to be

$$\nabla \phi \equiv 2 \frac{\partial \phi}{\partial \mathbf{x}^*} \quad (52)$$

and is normal to the surface of the cost function [35, Appendix B]. We similarly define the Hessian matrix  $\mathcal{H}\phi$  of  $\phi$  to be

$$\mathcal{H}\phi \equiv \frac{\partial}{\partial \mathbf{x}} [\nabla \phi]^T = 2 \frac{\partial}{\partial \mathbf{x}} \left[ \left( \frac{\partial \phi}{\partial \mathbf{x}^*} \right)^T \right]. \quad (53)$$

The cost functions that we encounter in (14) and (16) are of the form

$$\phi(\mathbf{x}) = \frac{1}{2\sigma_n^2} \|\mathbf{y} - H\mathbf{x}\|^2 + r(\mathbf{x}) = \frac{1}{2\sigma_n^2} (\mathbf{y}^H \mathbf{y} - \mathbf{y}^H H\mathbf{x} - \mathbf{x}^H H^H \mathbf{y} + \mathbf{x}^H H^H H\mathbf{x}) + r(\mathbf{x}) \quad (54)$$

where  $r(\mathbf{x})$  is the regularising constraint on  $\mathbf{x}$ . Differentiating gives the gradient vector and the Hessian matrix as

$$\nabla \phi = \sigma_n^{-2} (H^H H\mathbf{x} - H^H \mathbf{y}) + \nabla r, \quad (55)$$

$$\mathcal{H}\phi = \sigma_n^{-2} H^H H\mathbf{x} + \mathcal{H}r. \quad (56)$$

In the case of  $l^2$ -norm regularisation (14),  $r(\mathbf{x}) = \frac{1}{2} \mathbf{x}^H C_x^{-1} \mathbf{x}$  and  $\nabla r = C_x^{-1} \mathbf{x}$  and  $\mathcal{H}r = C_x^{-1}$ .

In the case of  $l^1$ -norm regularisation (16),  $r(\mathbf{x}) = \sqrt{3} \sigma_x^{-1} \sum_{i=1}^N |x_i| = \sqrt{3} \sigma_x^{-1} \sum_{i=1}^N \sqrt{x_i x_i^*}$  and the gradient

vector and Hessian matrix of  $r(\mathbf{x})$  are given component-wise by

$$[\nabla r]_i = 2 \times \frac{\sqrt{3}}{\sigma_x} \times \frac{\partial}{\partial x_i^*} \left( \sqrt{x_i x_i^*} \right) = \frac{\sqrt{3}}{\sigma_x} \cdot \frac{x_i}{|x_i|}, \quad (57)$$

$$[\mathcal{H}r]_{ik} = \begin{cases} 0 & \text{if } i \neq k, \\ \frac{\sqrt{3}}{\sigma_x} \times \frac{\partial}{\partial x_i} \left( \frac{x_i}{\sqrt{x_i x_i^*}} \right) = \frac{\sqrt{3}}{2\sigma_x} \cdot \frac{1}{|x_i|} & \text{if } i = k. \end{cases} \quad (58)$$

## D Correction of Artifacts in the In Vitro Results

In our *in vitro* experiments, we found that the restored images of the phantom with spherical inclusions exhibited artifacts in the form of lateral lines at equally spaced axial intervals (see Fig. 5). The regular axial spacing of these lines and the fact that they appeared in all of the restored images suggest that they are caused by the non-exact nature of our approximation to the blurring operator.

To correct these artifacts, we first observed that the background material in the phantom is supposed to be uniformly echogenic and, in the absence of time-gain compensation, should appear as pure speckle with the same echogenicity at all axial depths. Based on this observation, we performed the following steps to remove the lateral artifacts from each restored image:

1. We applied each restoration algorithm to an image consisting only of the background material. The output of this step is an image that exhibits the same lateral artifacts but is otherwise just pure speckle with smooth axial variations in background intensity due to time-gain compensation in the original background image.
2. We averaged the envelopes of the axial traces in the image from step (1) to yield a correction curve. This correction curve follows approximately the profile of the smooth axial variations in background intensity except for sharply-defined peaks at the locations of the artifacts. We applied a low-pass filter to the correction curve to smooth out irregularities.
3. We divided each axial trace in the restored image of the phantom by this correction curve to equalise the background intensity and hence cancel out the lateral artifacts.

Our artifact correction scheme is illustrated in Fig. 5. In addition to correcting the lateral artifacts introduced by restoration, this technique also has the useful side effect of cancelling out the smooth axial variations in background intensities due to time-gain compensation to more accurately reflect the uniform background echogenicity of the phantom (in the artifact-corrected restored images in Fig. 4, the background intensities appear uniform across all axial depths).

## References

- [1] O. Michailovich and D. Adam, "A novel approach to the 2-D blind deconvolution problem in medical ultrasound," *IEEE Trans. Med. Imag.*, vol. 24, no. 1, pp. 86–104, Jan. 2005.
- [2] S. Wan, B. I. Raju, and M. A. Srinivasan, "Robust deconvolution of high-frequency ultrasound images using higher-order spectral analysis and wavelets," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 50, no. 10, pp. 1286–1295, Oct. 2003.
- [3] T. Taxt and J. Strand, "Two-dimensional noise-robust blind deconvolution of ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, no. 4, pp. 861–866, July 2001.

- [4] T. Taxt, "Three-dimensional blind deconvolution of ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, no. 4, pp. 867–871, July 2001.
- [5] Y. Li, "Phase aberration correction using near-field signal redundancy - Part I: Principles," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 44, no. 2, pp. 355–371, Mar. 1997.
- [6] Y. Li, D. Robinson, and D. Carpenter, "Phase aberration correction using near-field signal redundancy - Part II: Experimental results," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 44, no. 2, pp. 372–379, Mar. 1997.
- [7] L. Nock, G. E. Trahey, and S. W. Smith, "Phase aberration correction in medical ultrasound using speckle brightness as a quality factor," *Journal of the Acoustical Society of America*, vol. 85, no. 5, pp. 1819–1833, May 1989.
- [8] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [9] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, pp. 17–33, 2003.
- [10] Y. Yue, M. M. Croitoru, A. Bidani, J. B. Zwischenberger, and J. W. Clark, Jr., "Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images," *IEEE Trans. Med. Imag.*, vol. 25, no. 3, pp. 297–311, 2006.
- [11] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 229–240, 2006.
- [12] O. V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 53, no. 1, pp. 64–78, 2006.
- [13] C. P. Loizou, C. S. Pattichis, C. I. Christodoulou, R. S. H. Istepanian, M. Pantziaris, and A. Nicolaides, "Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 52, no. 10, pp. 1653–1669, 2005.
- [14] N. Gupta, M. N. S. Swamy, and E. Plotkin, "Despeckling of medical ultrasound images using data and rate adaptive lossy compression," *IEEE Trans. Med. Imag.*, vol. 24, no. 6, pp. 743–754, 2005.
- [15] O. Husby, T. Lie, T. Lango, J. Hokland, and H. Rue, "Bayesian 2-D deconvolution: A model for diffuse ultrasound scattering," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, no. 1, pp. 121–130, Jan. 2001.
- [16] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [17] R. Neelamani, H. Choi, and R. Baraniuk, "ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 418–433, Feb. 2004.
- [18] J. Ng, R. Prager, N. Kingsbury, G. Treece, and A. Gee, "Modeling ultrasound imaging as a linear, shift-variant system," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 53, no. 3, pp. 549–563, Mar. 2006.
- [19] J. C. Bamber and R. J. Dickinson, "Ultrasonic B-scanning: A computer simulation," *Phys. Med. Biol.*, vol. 25, no. 3, pp. 463–479, 1980.
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, ser. Signal Processing. Prentice Hall, 1993, vol. 1.
- [21] J. G. Nagy and D. P. O'Leary, "Fast iterative image restoration with a spatially-varying PSF," in *Advanced Signal Processing Algorithms, Architectures, and Implementations IV*, F. T. Luk, Ed., vol. 3162, 1997, pp. 388–399.

- [22] —, “Restoring images degraded by spatially-variant blur,” *SIAM J. Sci. Comput.*, vol. 19, pp. 1063–1082, 1998.
- [23] F. Dellaert, “The expectation maximization algorithm,” Georgia Institute of Technology, Tech. Rep., Feb. 2002.
- [24] L. Sendur and I. W. Selesnick, “Bivariate shrinkage with local variance estimation,” *IEEE Signal Processing Lett.*, vol. 9, no. 12, pp. 438–441, Dec. 2002.
- [25] —, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Trans. Signal Processing*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.
- [26] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. V. der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. Philadelphia, PA: SIAM, 1994.
- [27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 2002.
- [28] K. M. Perry and S. J. Reeves, “A practical stopping rule for iterative signal restoration,” *IEEE Trans. Signal Processing*, vol. 42, no. 7, pp. 1829–1833, July 1994.
- [29] M. F. Hutchinson, “A stochastic estimator for the trace of the influence matrix for Laplacian smoothing splines,” *Comm. Statistics - Simulation Comput.*, vol. 19, no. 2, pp. 433–450, 1990.
- [30] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Processing Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
- [31] N. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” *Applied and Computational Harmonic Analysis*, vol. 10, pp. 234–253, 2001.
- [32] —, “Image processing with complex wavelets,” *Phil. Trans. R. Soc. Lond. A*, vol. 357, pp. 2543–2560, 1999.
- [33] J. A. Jensen, “Speed-accuracy trade-offs in computing spatial impulse responses for simulating medical ultrasound imaging,” *Journal of Computational Acoustics*, vol. 9, no. 3, pp. 731–744, 2001.
- [34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed., A. Jeffrey, Ed. Academic Press, 1994.
- [35] S. Haykin, *Adaptive Filter Theory*, 3rd ed., ser. Information and System Sciences. Prentice Hall, 1996.
- [36] D. H. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *IEE Proceedings H (Microwaves, Optics and Antennas)*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

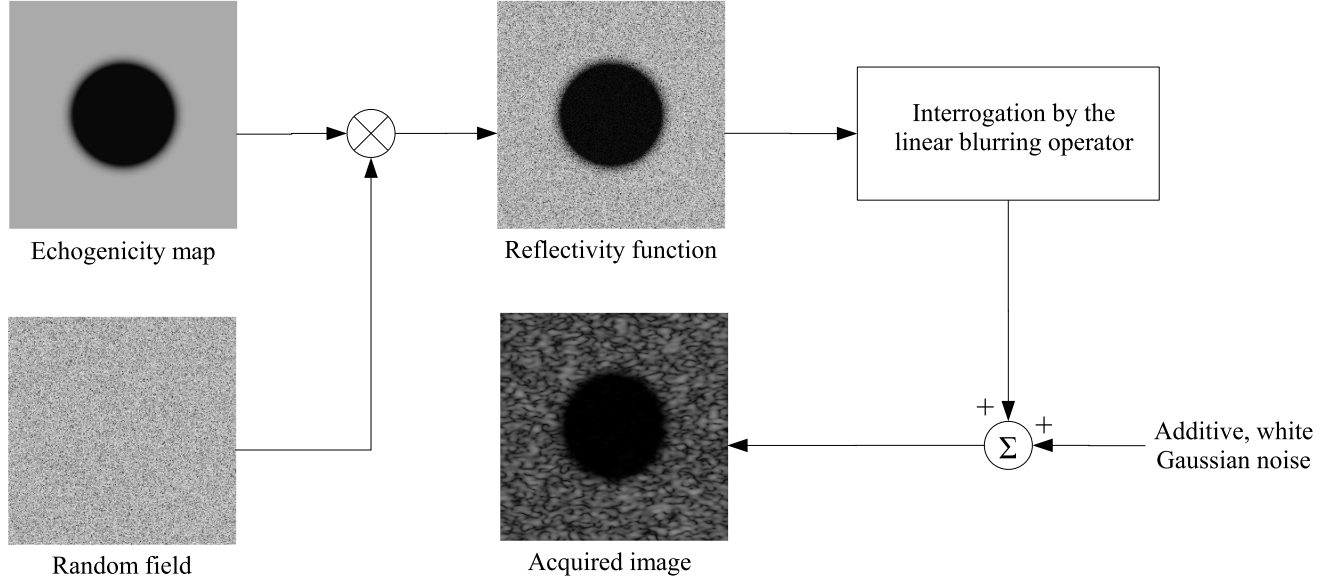
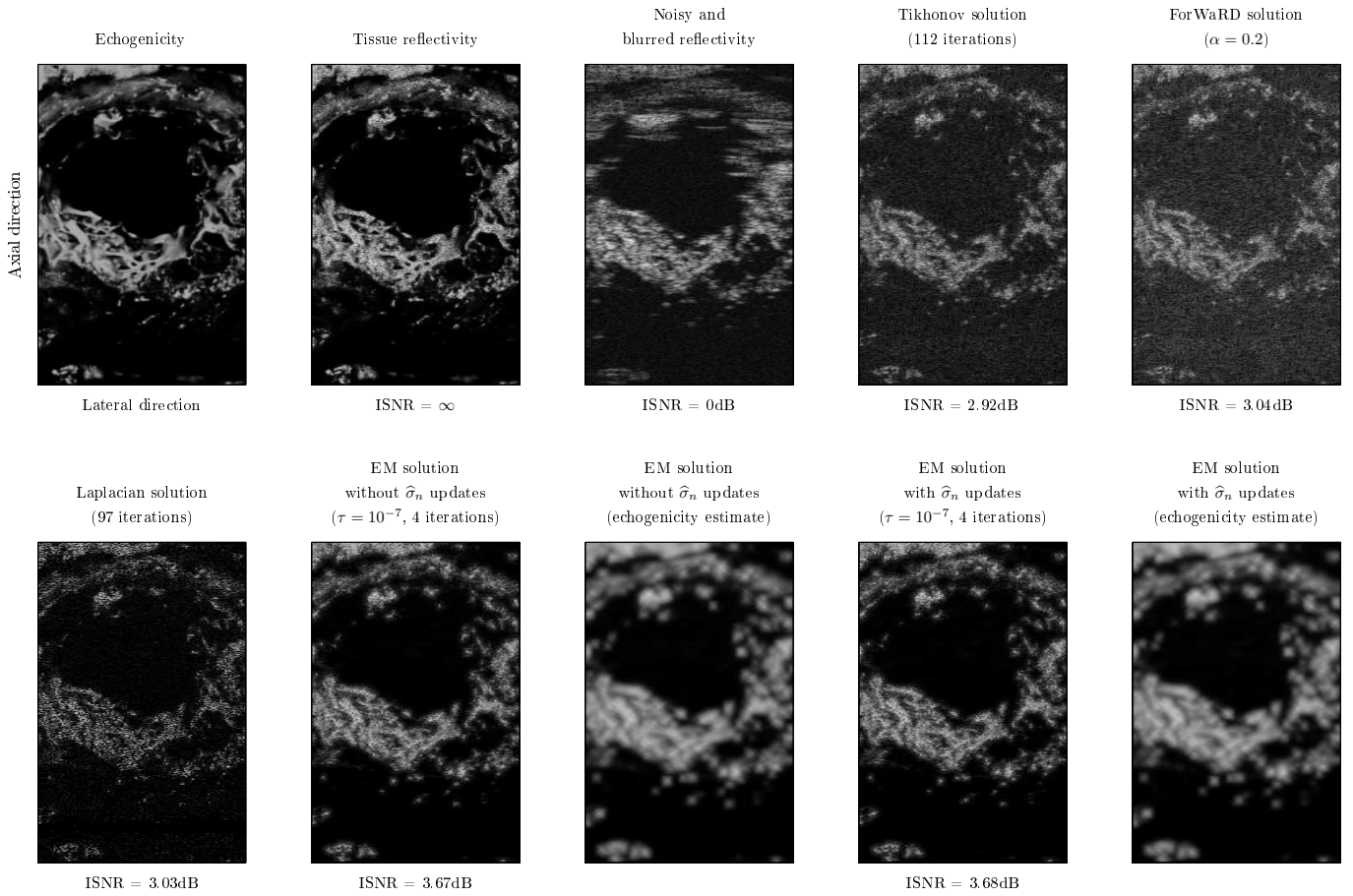


Figure 1: Simple illustration of our model of tissue reflectivity for a hypothetical anechoic cyst. The echogenicity map is smooth, except for the discontinuity at the boundary of the cyst. Multiplying this echogenicity map by a field of random, uncorrelated and identically distributed random variables yields a plausible reflectivity function which exhibits piecewise smoothness on a macroscopic scale and a random texture on a microscopic scale. Interrogation by the linear blurring operator and adding white Gaussian noise yields a typical speckled image. The grey-scale levels in all of the images represent logarithmically compressed amplitudes and span a dynamic range of 40dB.

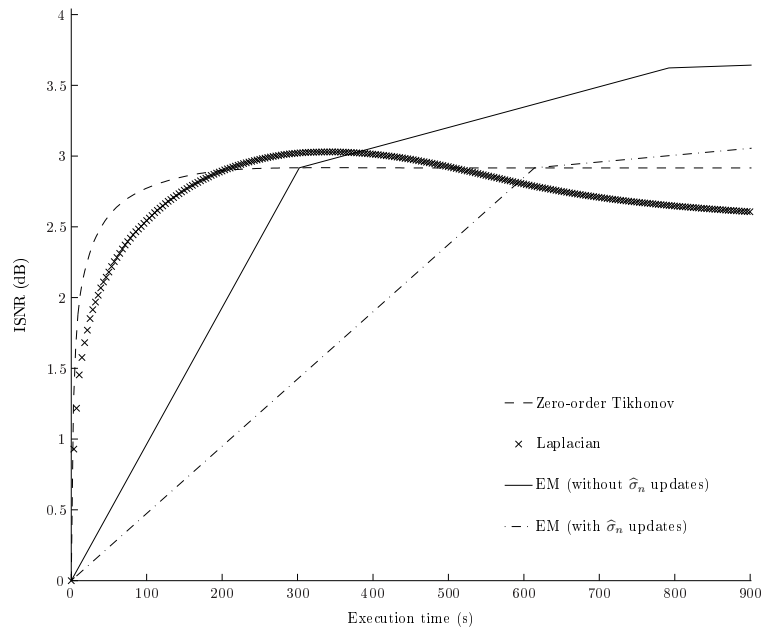
Table 1: List of symbols.

In this table,  $\mathbf{a} \in \mathbb{C}^N$ ,  $A \in \mathbb{C}^{N \times N}$ ,  $\phi: \mathbb{C}^N \rightarrow \mathbb{R}$  and  $p(\mathbf{a})$  is the probability density function of  $\mathbf{a}$ .

Symbol	Description	Mathematical definition
$(\mathbf{a})_n$	$n$ th element of $\mathbf{a}$	
$(A)_{mn}$	Element of $A$ in the $m$ th row and $n$ th column	
$A^T$	Non-conjugate transpose of $A$	
$A^H$	Conjugate (Hermitian) transpose of $A$	
$ A $	Determinant of $A$	
$\text{Tr}(A)$	Trace of $A$	$\sum_{n=1}^N (A)_{nn}$
$\ \mathbf{a}\ $	$l^2$ -norm of $\mathbf{a}$	$\sqrt{\mathbf{a}^H \mathbf{a}}$
$\ A\ _F$	Frobenius norm of $A$	$\sqrt{\sum_{m=1}^N \sum_{n=1}^N  (A)_{mn} ^2}$
$E(\mathbf{a})$	Expected value of $\mathbf{a}$	$\int \mathbf{a} p(\mathbf{a}) d\mathbf{a}$
$E[\phi(\mathbf{a})]$	Expected value of $\phi(\mathbf{a})$	$\int \phi(\mathbf{a}) p(\mathbf{a}) d\mathbf{a}$
$\text{Var}(\mathbf{a})$	Covariance matrix of $\mathbf{a}$	$\int (\mathbf{a} - \mu)(\mathbf{a} - \mu)^H p(\mathbf{a}) d\mathbf{a}$ , $\mu = E(\mathbf{a})$

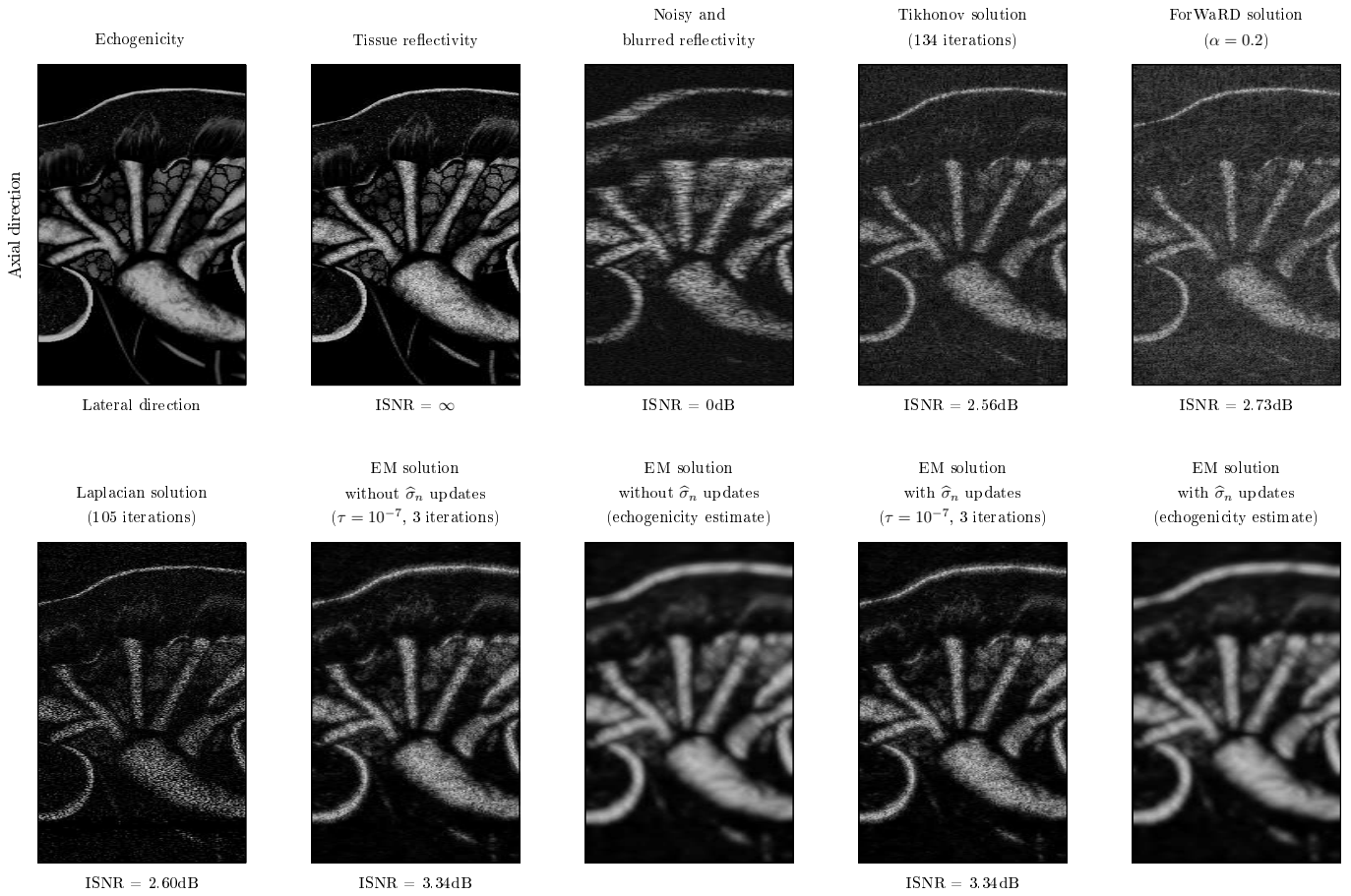


(a)

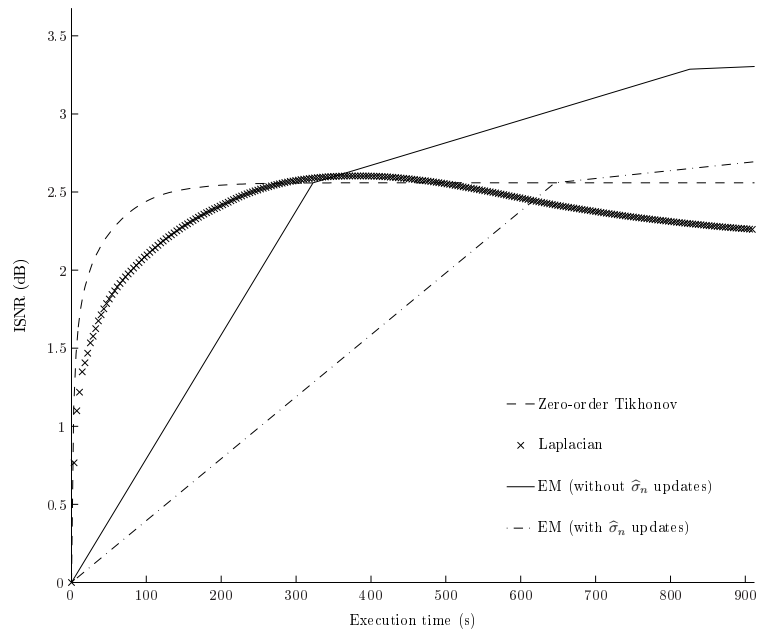


(b)

Figure 2: Simulation results for the heart image. (a) B-scan images of the true echogenicity, the true reflectivity, the corrupted reflectivity and the results of the various restoration schemes. The grey-scale levels represent logarithmically compressed amplitudes and span a dynamic range of 40dB. The lateral focus is approximately at the centre of the image. The echogenicity image was obtained with permission from [www.umdj.edu/pathweb/syspath/syslab\\_2/Slides\\_14/Slide\\_14\\_A/slide\\_14\\_a.htm](http://www.umdj.edu/pathweb/syspath/syslab_2/Slides_14/Slide_14_A/slide_14_a.htm). (b) Evolution of ISNR versus program execution time for the various restoration methods excluding ForWaRD. The execution times are for a personal computer running Matlab 7 with a 3.2GHz processor and 1GB of memory.



(a)



(b)

Figure 3: Simulation results for the kidney image. (a) B-scan images of the true echogenicity, the true reflectivity, the corrupted reflectivity and the results of the various restoration schemes. The grey-scale levels represent logarithmically compressed amplitudes and span a dynamic range of 40dB. The lateral focus is approximately in the centre of the image. The echogenicity image was obtained with permission from [www.med-ars.it/galleries/kydney\\_10.htm](http://www.med-ars.it/galleries/kydney_10.htm). (b) Evolution of ISNR versus program execution time for the various restoration methods excluding ForWaRD. The execution times are for a personal computer running Matlab 7 with a 3.2GHz processor and 1GB of memory.

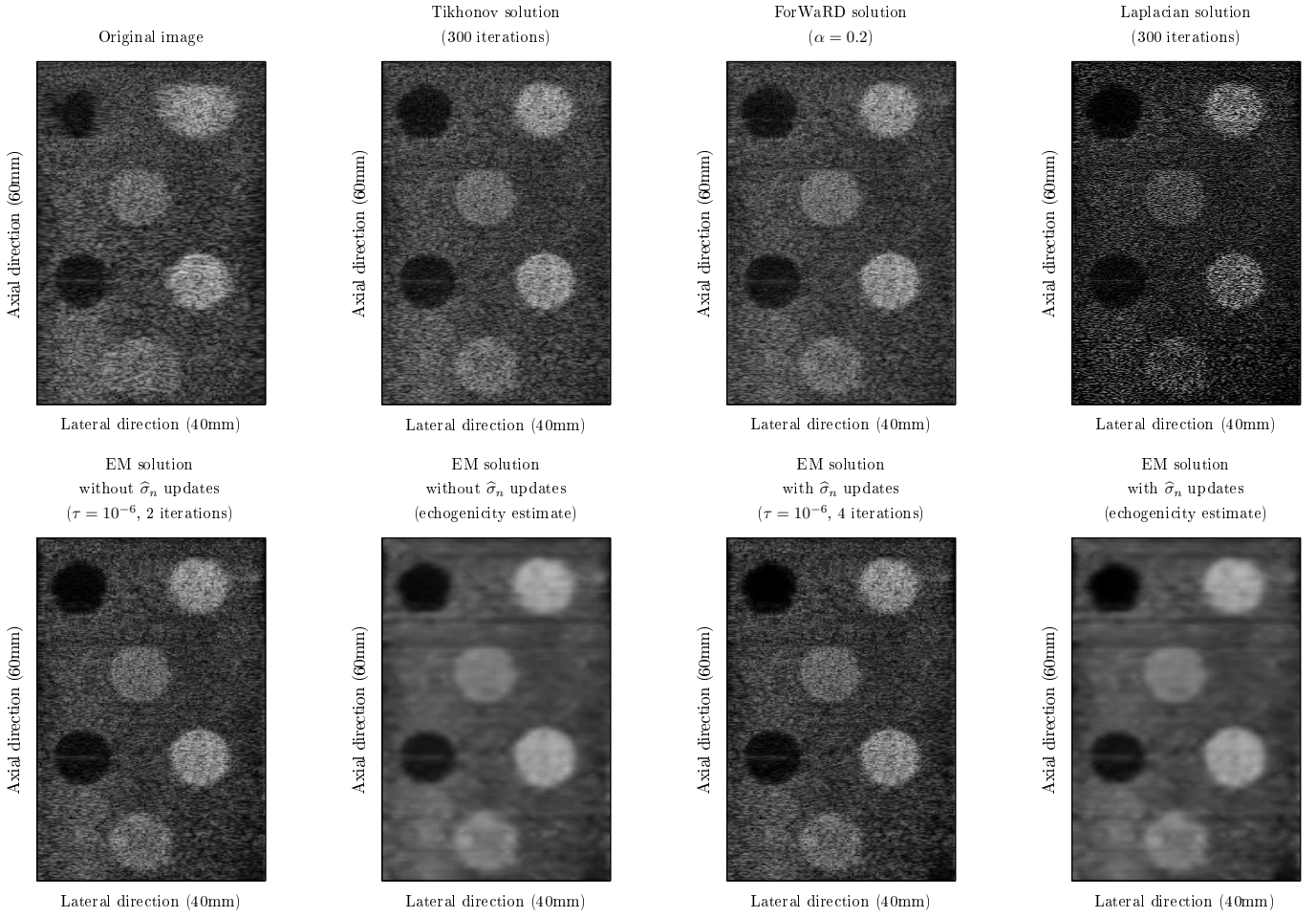
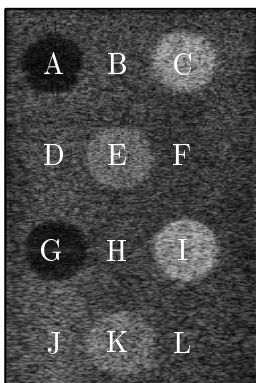


Figure 4: *In vitro* results for an image of a phantom containing spherical inclusions. The lateral focus is 26mm below the top of the image. The grey-scale levels represent logarithmically compressed amplitudes and span a dynamic range of 40dB.

Table 3: Local contrasts in the original and restored images of the phantom containing spherical inclusions. The image on the left labels 12 regions in the phantom and the table on the right contains contrast values (in dB) between selected pairs of adjacent  $6.25 \times 6.25$ mm regions. All contrast values were calculated from the restored images after artifact correction.



	Original image	Tikhonov solution	ForWaRD solution	Laplacian solution	EM solution (without $\hat{\sigma}_n$ updates)	EM solution (with $\hat{\sigma}_n$ updates)
B:A	6.81	9.42	9.28	14.01	15.69	21.14
C:B	5.83	5.94	5.79	6.49	6.32	6.57
E:D	2.49	2.75	2.87	2.69	2.89	2.93
F:E	-5.28	-5.23	-5.73	-5.26	-5.60	-5.71
H:G	6.94	6.72	7.07	6.94	8.77	9.56
I:H	6.52	6.42	6.45	6.73	7.00	7.18
K:J	1.00	1.33	1.41	1.57	1.71	1.65
L:K	-3.85	-4.21	-4.37	-4.79	-5.58	-5.44

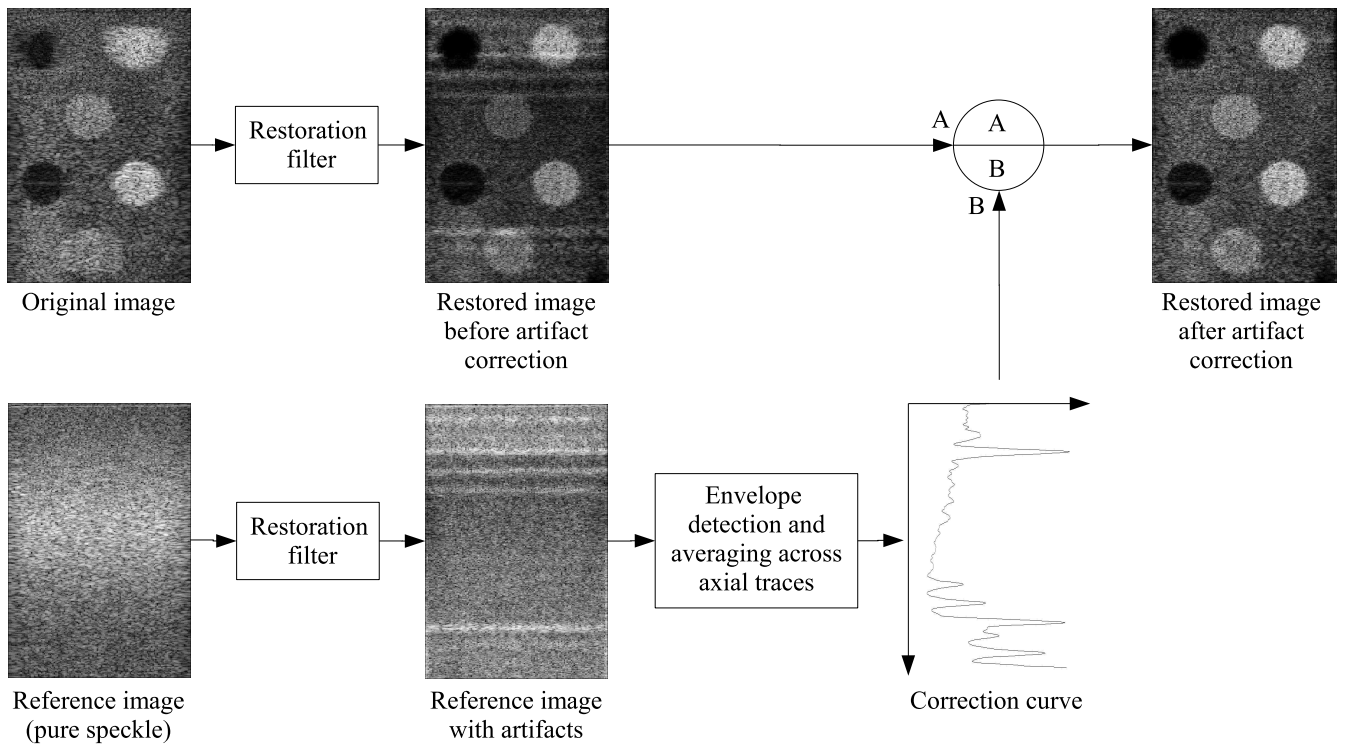


Figure 5: Correction of artifacts introduced by errors in the approximation to the blurring operator and exacerbated by image restoration. Applying the restoration filter to a reference image of uniformly echogenic speckle produces an image that exhibits artifacts from the blurring operator approximation. Envelope detection of this filtered reference image followed by averaging across its axial traces yields a correction curve. Dividing each axial trace in the restored image of the phantom by the correction curve removes the artifacts. Each image in this illustration has been logarithmically compressed and scaled to a dynamic range of 40dB.