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

James Ng, Richard Prager, Nick Kingsbury, *Member, IEEE*, Graham Treece, and Andrew Gee

***Abstract*—The clinical utility of pulse-echo ultrasound images is severely limited by inherent poor resolution that 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 modeling of the underlying tissue reflectivity that is imaged. The acoustic response of soft biological tissues has statistics that differ substantially from the natural images considered in mainstream image processing: although, 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 pseudo-random texture (manifested in the amplitude image as speckle) due to the dense concentrations of small, weakly scattering particles. Recognizing that this pseudo-random texture is diagnostically important for tissue identification, we propose modeling 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-maximization (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 that have better image quality and greater fidelity to the true tissue reflectivity than other restoration techniques based on simpler regularizing constraints.**

I. 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 nontrivial 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 that can adequately capture significant image features without being computationally prohibitive. 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-maximization (EM) algorithm to yield a computationally efficient iterative restoration method based on Wiener filtering and wavelet denoising. Our algorithm implicitly assumes the blurring operator to be known *a priori*, and in our experiments, we approximated the blurring operator by point-spread functions simulated in Field II [1]. In practice, dispersive attenuation and phase aberrations can cause the blurring operator to deviate from what is theoretically predicted; however, these deviations can be accommodated to some degree by estimating the blurring operator directly from the acquired image with blind deconvolution methods (e.g., [2]–[5]), and there exist also a number of ways to correct for phase aberrations (e.g., [6]–[8]).

Wavelet transforms have been very successful in the processing of so-called natural images, i.e., images that 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 computationally efficient wavelet-based algorithms for image compression and analysis (see, for example, [9], [10]). 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 that are imaged by pulse-echo ultrasound do not exhibit piecewise-smoothness in the same way as the natural images considered in mainstream image processing: although, 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 behavior that manifests itself as the characteristic speckle pattern seen in ultrasound images.

Speckle obscures significant image features and degrades the resolvability of structures. 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 [11]–[15]). However, despite its negative effect on image quality, speckle also contains useful textural in-

Manuscript received May 17, 2006; accepted September 8, 2006.
The authors are with the Department of Engineering, University of Cambridge, Cambridge, UK (e-mail: jkhn2@cam.ac.uk).
Digital Object Identifier 10.1109/TUFFC.2007.278

formation that can assist in the identification of tissue type. Hence, we believe that restoration methods for pulse-echo ultrasound aimed at improving image quality for diagnostic purposes are required, not only to enhance image resolution, but also to preserve the textural information inherently present in speckle. Nevertheless, we recognize 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 modeling 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 behaviors separately. Furthermore, we observe that, because 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 modeling the echogenicity and the pseudo-random texture separately is not new and was proposed by Husby *et al.* [16] who modeled 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, nonlinear shrinkage rules that 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 in-phase/quadrature data that retain the same phase information but can be sampled at lower rates (this, in turn, further reduces the computational load).

We acknowledge the work of Figueiredo and Nowak [17], who developed an iterative algorithm based on EM for the restoration of blurred natural images that also alternate between Wiener filtering and wavelet-domain denoising. Their work, in turn, may be viewed as an extension of Neelamani *et al.*'s [18], [19] Fourier-wavelet regularized deconvolution (ForWaRD) method that applies an underregularized Wiener filter to the blurred image and follows this up with wavelet denoising. We emphasize, however, that, although there is a procedural similarity between our algorithm and these others, our algorithm differs from these others conceptually and was developed for tissue reflectivities with image statistics that are significantly different from those of natural images.

In the sections that follow, we first present a model for weak, linear scattering and describe a method for approximating the global shift-variant blurring operator by a collection of locally shift-invariant blurring operators. We then present our model of tissue reflectivity and show how EM can be applied to this model 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 I.

II. Background

The linear model for pulse-echo ultrasound imaging, based on the assumption of weak scattering and application of the first Born approximation, states that the RF image may be modeled as the result of applying a linear blurring operator to a three-dimensional tissue reflectivity. This three-dimensional model may be reduced to two dimensions if we assume that the acoustic properties of the interrogated tissue are approximately uniform over the support of the blurring kernel in the direction perpendicular to the scanning plane. To allow for sampling at a lower rate, the RF image may be demodulated to baseband, and because demodulation is a linear operation, an analogous linear model also holds at baseband.

In practice, processing of the ultrasound image is carried out on digitized 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, and $\mathbf{n}$ to be an $N \times 1$ noise vector that accounts for measurement error. The discrete approximation to the linear blurring model is then given by:

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}. \tag{1}$$

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]. Defining its covariance matrix $\mathbf{E}(\mathbf{n}\mathbf{n}^{\mathrm{H}}) = 2\sigma_n^2 I_N$, and with a slight abuse of notation, we may write the probability density function of $\mathbf{n}$ as:

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

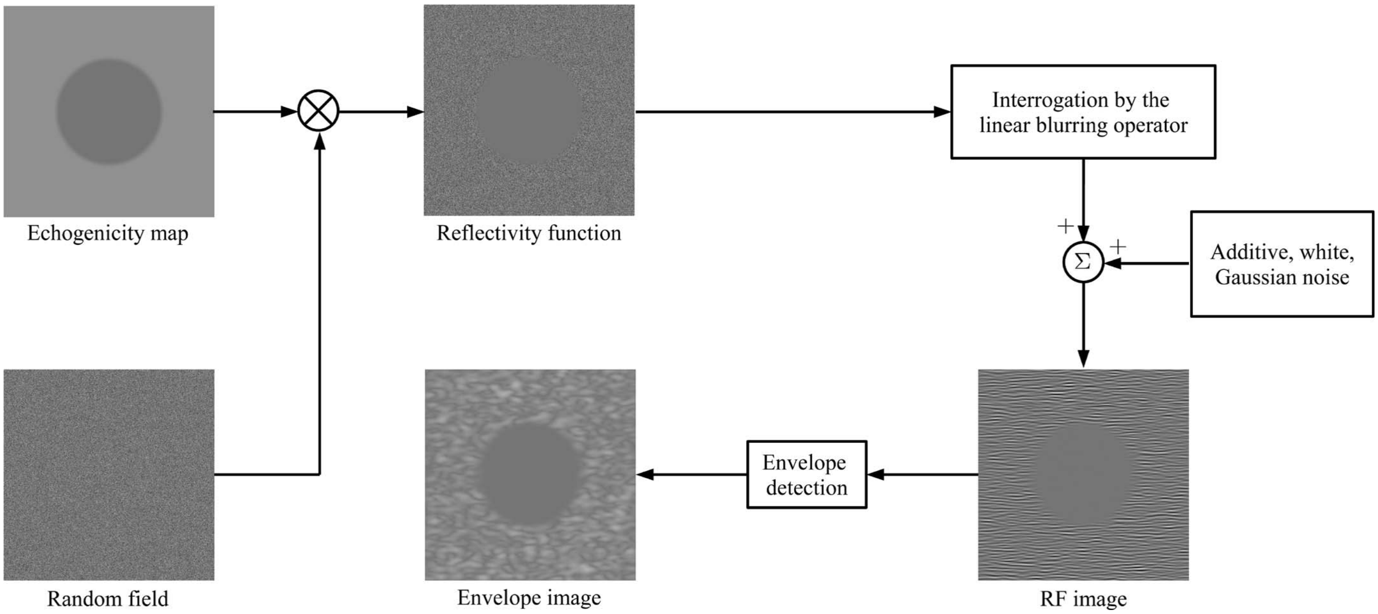


Fig. 1. Simple illustration of our model of tissue reflectivity for a hypothetical anechoic cyst. The echogenicity map is smooth, except for the discontinuity at the boundary of the cyst. Multiplying this echogenicity map by a field of random, uncorrelated, and identically distributed random variables yields a plausible reflectivity function that 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.

TABLE I
List of Symbols.[1]

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

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

where σ_n is necessarily real valued. The real and imaginary parts of each element of $\mathbf{n}$ are uncorrelated and have a variance of σ_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 in which the blurring operator is shift-invariant and can be characterized by a single point-spread function, a product of the form $H\mathbf{x}$ is just a discrete convolution of $\mathbf{x}$ with the point-spread function and can be efficiently computed by multiplying together the discrete Fourier transform (DFT) coefficients of $\mathbf{x}$ and the DFT coefficients of the point-spread function and taking the inverse DFT of the product. Likewise, a product of the form $H^{\mathrm{H}}\mathbf{x}$ is a discrete convolution of $\mathbf{x}$ with the complex conjugate of the spatially reversed point-spread function and can be computed by a similar multiplication in the DFT domain with the DFT coefficients of the point-spread function replaced by their complex conjugates.

Unfortunately, in pulse-echo ultrasound imaging, the width of the acoustic beam interrogating the subject is nonuniform and varies with axial distance, giving rise to a blurring operator that is shift-variant. Michailovich and Adam [2] proposed handling this shift-variant blurring operator with Nagy and O'Leary's [21], [22] method of partitioning image space into a number of regions within which the blurring operator is assumed to be locally shift-invariant. By defining a point-spread function within each of these approximately shift-invariant regions, we can approximate the global blurring operator by convolving each region with its point-spread function, weighting the results, and summing them. Mathematically, we may write:

$$H \approx \sum_n D_n H_n, \tag{3}$$

where H_n is the shift-invariant blurring operator corresponding to the nth 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:

$$H^{\mathrm{H}} \approx \sum_n H_n^{\mathrm{H}} D_n, \tag{4}$$

which is equivalent to weighting each shift-invariant region, convolving each weighted region with the complex conjugate of its spatially reversed point-spread function 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.

III. Maximum A Posteriori 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 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 regularizing constraint that enforces feasibility by constraining our estimate of $\mathbf{x}$ to belong to a predefined functional subspace. This usually leads to the optimization of a cost function that trades off fidelity to the observed image $\mathbf{y}$ against the regularizing 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 that realization of $\mathbf{x}$ which maximizes its posterior probability $p(\mathbf{x} \mid \mathbf{y}, \sigma_n)$. Bayes's rule states that $p(\mathbf{x} \mid \mathbf{y}, \sigma_n) \propto p(\mathbf{y} \mid \mathbf{x}, \sigma_n)\, p(\mathbf{x} \mid \sigma_n)$, and because $\mathbf{x}$ and σ_n are independent, $p(\mathbf{x} \mid \sigma_n) = p(\mathbf{x})$. Taking logarithms, our MAP estimate $\hat{\mathbf{x}}$ may be written as:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \left[\ln p(\mathbf{y} \mid \mathbf{x}, \sigma_n) + \ln p(\mathbf{x})\right], \tag{5}$$

where the log-likelihood $\ln p(\mathbf{y} \mid \mathbf{x}, \sigma_n)$ enforces fidelity to $\mathbf{y}$, and the log-prior $\ln p(\mathbf{x})$ is the regularizing 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 (2) to give the likelihood distribution:

$$p(\mathbf{y} \mid \mathbf{x}, \sigma_n) = -\frac{1}{(2\pi\sigma_n^2)^N} \exp\left(-\frac{1}{2\sigma_n^2}\|\mathbf{y} - H\mathbf{x}\|^2\right), \tag{6}$$

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

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left[\frac{1}{2\sigma_n^2}\|\mathbf{y} - H\mathbf{x}\|^2 - \ln p(\mathbf{x})\right]. \tag{7}$$

A. l^2-Norm Regularization

One of the more common methods of regularization is to constrain the weighted l^2-norm of $\mathbf{x}$, a technique known as Tikhonov regularization [23]. Constraining the weighted l^2-norm of $\mathbf{x}$ is equivalent to modeling $\mathbf{x}$ as a Gaussian random vector with zero mean and covariance matrix $\mathbf{E}(\mathbf{x}\mathbf{x}^{\mathrm{H}}) = 2C_x$. We may then write $p(\mathbf{x}) \propto \exp(-\frac{1}{2}\mathbf{x}^{\mathrm{H}} C_x^{-1}\mathbf{x})$ and:

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \left(\frac{1}{2\sigma_n^2}\|\mathbf{y} - H\mathbf{x}\|^2 + \frac{1}{2}\mathbf{x}^{\mathrm{H}} C_x^{-1}\mathbf{x}\right) \\ &= \left(H^{\mathrm{H}} H + \sigma_n^2 C_x^{-1}\right)^{-1} H^{\mathrm{H}}\mathbf{y}.\end{aligned} \tag{8}$$

The closed-form expression on the rightmost side of (8) is the well-known Wiener filter, and its application to pulse-echo ultrasound images was proposed by Taxt and Strand [4] and Taxt [5] and by Michailovich and Adam [2]. The special case $C_x = \sigma_x^2 I_N$, where σ_x is real-valued, is referred to as zero-order Tikhonov regularization.

The ForWaRD method, which was briefly mentioned in Section I, is a nonlinear, wavelet-based extension of zero-order Tikhonov regularization: it applies an under-regularized Wiener filter (formed by replacing σ_n^2 with $\alpha\sigma_n^2$ where $\alpha \in (0, 1]$ is an under-regularization parameter) to $\mathbf{y}$ and then follows this up with wavelet shrinkage to remove the noise amplified by the under-regularization. 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 [3].

B. l^1-Norm Regularization

Michailovich and Adam [2] proposed modeling 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 permit better recovery of the stronger reflectors at structural boundaries. Their definition of the Laplacian distribution in [2] 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 σ_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:

$$p(x_i) \propto \exp\left(-\frac{\sqrt{3}}{\sigma_x}|x_i|\right). \tag{9}$$

This probability density function is circularly symmetric in the complex plane, which keeps the phase of each x_i uniformly distributed between 0 and 2π.

Assuming the components of $\mathbf{x}$ to be independent and identically distributed, the probability density function in (9) gives rise to a regularizing constraint on the l^1-norm instead of the l^2-norm, and our MAP estimation becomes:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left(\frac{1}{2\sigma_n^2}\|\mathbf{y} - H\mathbf{x}\|^2 + \frac{\sqrt{3}}{\sigma_x}\sum_{i=1}^{N}|x_i|\right). \tag{10}$$

IV. Expectation-Maximization

A. 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 regularizing constraints of (8) and (10). The reflectivities of biological tissues typically exhibit nonstationary 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 behavior 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:

$$\mathbf{x} = S\,\mathbf{w}, \tag{11}$$

and we model $\mathbf{w}$ as a complex-valued, zero-mean Gaussian vector with covariance matrix $\mathbf{E}(\mathbf{w}\mathbf{w}^{\mathrm{H}}) = 2I_N$ (this is consistent with the model in [16]). Because 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:

$$p(\mathbf{x} \mid S) \propto \frac{1}{|S^2|}\exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{H}}S^{-2}\mathbf{x}\right). \tag{12}$$

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 $\mathbf{E}(\mathbf{x}\mathbf{x}^{\mathrm{H}}) = 2S^2$, and an estimate of $\mathbf{x}$ could be calculated from $\mathbf{y}$ by applying the Wiener filter as given in (8).
- 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.

B. The EM Algorithm

The EM algorithm is an iterative procedure for MAP estimation in problems in which the joint likelihood of some observed data and a set of unobservable (or hidden parameters) is specified but not the marginal likelihood of the observed data. It provides an alternative to explicit marginalization of the joint likelihood that is often not analytically tractable. Each iteration of EM yields an estimate of the parameter of interest, which increases its posterior and convergence to a local maximum of the posterior distribution is guaranteed.

We present, without proof, the mechanics of the EM algorithm for MAP estimation as described in [24]. Let $\mathbf{\Theta}$ be the parameter of interest, let $\mathbf{U}$ be the observed data from which we wish to estimate $\mathbf{\Theta}$, let $\mathbf{J}$ be the nuisance parameter, and let $\widehat{\mathbf{\Theta}}_k$ be the estimate of $\mathbf{\Theta}$ at the kth iteration. At the kth iteration, we execute:

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

$$\begin{aligned} Q(\mathbf{\Theta} \mid \widehat{\mathbf{\Theta}}_k) &= \mathbf{E}\left[\ln p(\mathbf{U}, \mathbf{J} \mid \mathbf{\Theta}) \mid \mathbf{U}, \widehat{\mathbf{\Theta}}_k\right] \\ &= \int p\left(\mathbf{J} \mid \mathbf{U}, \widehat{\mathbf{\Theta}}_k\right) \ln p(\mathbf{U}, \mathbf{J} \mid \mathbf{\Theta})\, d\mathbf{J}. \end{aligned} \tag{13}$$

- The M-step: We calculate the next estimate of $\mathbf{\Theta}$:

$$\widehat{\mathbf{\Theta}}_{k+1} = \arg\max_{\mathbf{\Theta}} \left[Q\left(\mathbf{\Theta} \mid \widehat{\mathbf{\Theta}}_k\right) + \ln p(\mathbf{\Theta})\right]. \tag{14}$$

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

So far, we have implicitly assumed the variance $2\sigma_n^2$ of the additive noise term in (1) to be known *a priori*. Although 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 σ_n as a parameter of interest to be estimated. We introduce a flat prior for σ_n that corresponds to solving for its maximum likelihood (ML) estimate.

C. E-Step: Wiener Filtering

We define $\widehat{S}_k$ and $\hat{\sigma}_{n,k}$ to be the estimates of S and σ_n, respectively, at the kth iteration. Substituting $\mathbf{U} = \mathbf{y}$, $\mathbf{J} = \mathbf{x}$, and $\mathbf{\Theta} = \{S, \sigma_n\}$ into (13), we obtain:

$$Q\left(S, \sigma_n \mid \widehat{S}_k, \hat{\sigma}_{n,k}\right) = \int p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) \ln p(\mathbf{y}, \mathbf{x} \mid S, \sigma_n)\, d\mathbf{x}. \quad (15)$$

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

$$p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) = \frac{p\left(\mathbf{y} \mid \mathbf{x}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) p\left(\mathbf{x} \mid \widehat{S}_k, \hat{\sigma}_{n,k}\right)}{p\left(\mathbf{y} \mid \widehat{S}_k, \hat{\sigma}_{n,k}\right)} = \frac{p\left(\mathbf{y} \mid \mathbf{x}, \hat{\sigma}_{n,k}\right) p\left(\mathbf{x} \mid \widehat{S}_k\right)}{p\left(\mathbf{y} \mid \widehat{S}_k, \hat{\sigma}_{n,k}\right)}. \quad (16)$$

The denominator in the rightmost term of (16) does not depend on $\mathbf{x}$ and may be regarded as just a normalization constant. Substituting (6) and (12) into the numerator and simplifying, we obtain:

$$p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) \propto \exp \cdot \left\{-\frac{1}{2\hat{\sigma}_{n,k}^2}\left[\mathbf{x}^{\mathrm{H}}\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)\mathbf{x} - 2\,\mathrm{Re}\,\left(\mathbf{x}^{\mathrm{H}}H^{\mathrm{H}}\mathbf{y}\right)\right]\right\}, \quad (17)$$

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

$$C_k = \mathrm{Var}\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) = 2\hat{\sigma}_{n,k}^2\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^{-1}, \quad (18)$$

$$\mathbf{m}_k = \mathrm{E}\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) = \left(2\hat{\sigma}_{n,k}^2\right)^{-1} C_k H^{\mathrm{H}}\mathbf{y} = \left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^{-1} H^{\mathrm{H}}\mathbf{y}. \quad (19)$$

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 kth iteration. Comparing the rightmost expression in (19) with (8) 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 (15), 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:

$$Q = Q_1 + Q_2, \quad (20)$$

$$Q_1 = \int p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) \ln p(\mathbf{x} \mid S)\, d\mathbf{x}, \quad (21)$$

$$Q_2 = \int p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) \ln p(\mathbf{y} \mid \mathbf{x}, \sigma_n) d\mathbf{x}. \quad (22)$$

To expand Q_1, we substitute (12) into (21) and discard constants to yield:

$$\widetilde{Q}_1 = -2\ln|S| - \frac{1}{2}\int \mathbf{x}^{\mathrm{H}} S^{-2}\mathbf{x}\, p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) d\mathbf{x}. \quad (23)$$

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

$$\begin{aligned}\mathrm{E}\left(\mathbf{x}^{\mathrm{H}} S^{-2}\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) &= \mathrm{E}\left[\mathrm{Tr}\left(S^{-2}\mathbf{x}\mathbf{x}^{\mathrm{H}}\right) \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right] \\ &= \mathrm{Tr}\left[S^{-2}\mathrm{E}\left(\mathbf{x}\mathbf{x}^{\mathrm{H}} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right)\right] \\ &= \mathrm{Tr}\left[S^{-2}\left(C_k + \mathbf{m}_k\mathbf{m}_k^{\mathrm{H}}\right)\right] \\ &= \mathrm{Tr}\left(S^{-2}C_k\right) + \mathbf{m}_k^{\mathrm{H}} S^{-2}\mathbf{m}_k, \end{aligned} \quad (24)$$

and substituting this expression back into (23), we obtain:

$$\widetilde{Q}_1 = -2\ln|S| - \frac{1}{2}\left[\mathrm{Tr}\left(S^{-2}C_k\right) + \|S^{-1}\mathbf{m}_k\|^2\right]. \quad (25)$$

To expand Q_2, we substitute (6) into (22) and discard constants to yield:

$$\widetilde{Q}_2 = -2N\ln\sigma_n - \frac{1}{2\sigma_n^2}\int \|\mathbf{y} - H\mathbf{x}\|^2 p\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) d\mathbf{x}. \quad (26)$$

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

$$\begin{aligned}\mathrm{E}\left(\|\mathbf{y} - H\mathbf{x}\|^2 \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right) &= \mathbf{y}^{\mathrm{H}}\mathbf{y} - 2\mathrm{Re}\left[\mathbf{y}^{\mathrm{H}} H\, \mathrm{E}\left(\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right)\right] \\ &\quad + \mathrm{E}\left(\mathbf{x}^{\mathrm{H}} H^{\mathrm{H}} H\mathbf{x} \mid \mathbf{y}, \widehat{S}_k, \hat{\sigma}_{n,k}\right), \end{aligned} \quad (27)$$

and, evaluating each of these expectations, we obtain:

$$\begin{aligned}\widetilde{Q}_2 &= -2N\ln\sigma_n - \frac{1}{2\sigma_n^2}\Big[\mathbf{y}^{\mathrm{H}}\mathbf{y} - 2\mathrm{Re}\left(\mathbf{y}^{\mathrm{H}} H\mathbf{m}_k\right) \\ &\quad + \mathbf{m}_k^{\mathrm{H}} H^{\mathrm{H}} H\mathbf{m}_k + \mathrm{Tr}\left(H^{\mathrm{H}} H\, C_k\right)\Big] \\ &= -2N\ln\sigma_n - \frac{1}{2\sigma_n^2}\Big[\|\mathbf{y} - H\mathbf{m}_k\|^2 + \mathrm{Tr}\left(H^{\mathrm{H}} H\, C_k\right)\Big]. \end{aligned} \quad (28)$$

We conclude our discussion on the E-step with the important observation that, because our joint log-likelihood Q is the sum of a function purely of S and a function purely of σ_n, and because S is independent of σ_n, we can split our M-step into two separate optimization problems:

$$\begin{aligned}\widehat{S}_{k+1} &= \arg\max_S \left[\widetilde{Q}_1 + \ln p(S)\right] \\ &= \arg\max_S \left\{-2\ln|S| - \frac{1}{2}\Big[\mathrm{Tr}\left(S^{-2}C_k\right) \right. \\ &\qquad \left. + \|S^{-1}\mathbf{m}_k\|^2\Big] + \ln p(S)\right\}, \end{aligned} \tag{29}$$

$$\begin{aligned}\hat{\sigma}_{n,k+1} &= \arg\max_{\sigma_n} \widetilde{Q}_2 \\ &= \arg\max_{\sigma_n} \left\{-2N\ln\sigma_n - \frac{1}{2\sigma_n^2}\Big[\|\mathbf{y} - H\mathbf{m}_k\|^2 \right. \\ &\qquad \left. + \mathrm{Tr}\left(H^{\mathrm{H}} H\, C_k\right)\Big]\right\}, \end{aligned} \tag{30}$$

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

D. 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 (11) as multiplicative noise to be removed. We define $x_i = (\mathbf{x})_i$, $s_i = (S)_i$, and $w_i = (\mathbf{w})_i$, and we rewrite (11) componentwise as $x_i = s_i w_i$ (recall that x_i and w_i are complex-valued and 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 $\tilde{x}_i = \ln|x_i|$, $\tilde{s}_i = \ln s_i$ and $\tilde{w}_i = \ln|w_i|$, we obtain:

$$\tilde{x}_i = \tilde{s}_i + \tilde{w}_i, \quad i = 1, \ldots, N. \tag{31}$$

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

$$\begin{aligned} p(\tilde{w}_i) &= \exp\left[2\tilde{w}_i - \frac{1}{2}\exp(2\tilde{w}_i)\right], \\ \mathrm{E}(\tilde{w}_i) &= \frac{1}{2}(\ln 2 - \gamma) \approx 0.0580, \\ \mathrm{Var}(\tilde{w}_i) &= \frac{\pi^2}{24} \approx 0.4112, \end{aligned} \tag{32}$$

where γ 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.

Because the echogenicity is piecewise-smooth, we expect the log-echogenicity $\{\tilde{s}_i\}$ to also be piecewise-smooth and to have a sparse representation in the wavelet domain. We can therefore denoise by applying wavelet shrinkage: because most of the energy of $\{\tilde{s}_i\}$ will be concentrated into just a few wavelet coefficients, we can modify the wavelet coefficients of $\{\tilde{x}_i\}$ according to some shrinkage rule that attenuates the logarithmic noise term $\{\tilde{w}_i\}$. 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 (32). In practice, however, we have found that, because of the band-limitedness of each wavelet subband, the central limit theorem keeps the wavelet coefficients of $\{\tilde{w}_i\}$ approximately Gaussian. The additive noise is also usually assumed to have zero mean, so we need to subtract 1/2 $(\ln 2 - \gamma)$ from $\{\tilde{x}_i\}$ 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 [25], [26]. Hence, we may regard the logarithmic denoising of $\{\tilde{x}_i\}$ to recover $\{\tilde{s}_i\}$ and estimate $\{s_i\}$ as being equivalent to the MAP estimation problem:

$$\begin{aligned}\widehat{S} &= \arg\max_S p\left(S \mid \mathbf{x}\right) = \arg\max_S p\left(\mathbf{x} \mid S\right) p(S) \\ &= \arg\max_S \left[\ln p\left(\mathbf{x} \mid S\right) + \ln p(S)\right] \\ &= \arg\max_S \left[-2\ln|S| - \frac{1}{2}\mathbf{x}^{\mathrm{H}} S^{-2}\mathbf{x} + \ln p(S)\right] \\ &= \arg\max_S \left[-\sum_{i=1}^{N}\left(2\ln s_i + \frac{|x_i|^2}{2s_i^2}\right) + \ln p(S)\right], \end{aligned} \tag{33}$$

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

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 (29) as:

$$\widehat{S}_{k+1} = \arg\max_S \left[-\sum_{i=1}^{N}\left(2\ln s_i + \frac{|m_{i,k}|^2 + \sigma_{i,k}^2}{2s_i^2}\right) + \ln p(S)\right]. \tag{34}$$

If we now compare this form of the update rule with (33), we see that the two expressions are identical simply by letting $x_i^2 = |m_{i,k}|^2 + \sigma_{i,k}^2$. 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 regularization.

E. M-Step: Updating the Additive Noise Variance Estimate

Differentiating the log-likelihood $\widetilde{Q}_2$ in (28) yields:

$$\frac{d\widetilde{Q}_2}{d\sigma_n} = -\frac{2N}{\sigma_n} + \frac{1}{\sigma_n^3}\left[\mathrm{Tr}\left(H^{\mathrm{H}} H\, C_k\right) + \|\mathbf{y} - H\mathbf{m}_k\|^2\right], \tag{35}$$

```
Estimate σ̂_n² with the robust wavelet estimator.
Initialize Ŝ = 2N σ̂_n² ||y||^{-2} I_N.
While termination condition is not satisfied,
    E-step:
    Define C = 2σ̂_n² (H^H H + σ̂_n² Ŝ^{-2})^{-1}    (18)
    Calculate m = (2σ̂_n²)^{-1} C H^H y    (19)
    Calculate σ_i² = (C)_ii for i = 1,...,N.
    M-step:
    Calculate x̃_i = ln( sqrt(|m_i|² + σ_i²) ) = (1/2) ln(|m_i|² + σ_i²) for i = 1,...,N.
    Estimate {s̃_i} by applying wavelet shrinkage to {x̃_i + (1/2)(γ − ln 2)}.
    Calculate ŝ_i = exp(s̃_i) for i = 1,...,N.
    Calculate σ̂_n² = (2N)^{-1} [||y − Hm||² + Tr(H^H H C)]    (36).
end
```

and setting the derivative to zero, we obtain:

$$\hat{\sigma}_{n,k+1}^2 = \frac{1}{2N}\left[\mathrm{Tr}\left(H^{\mathrm{H}} H\, C_k\right) + \|\mathbf{y} - H\mathbf{m}_k\|^2\right]. \tag{36}$$

The second derivative of the cost function at this value of σ_n is given by:

$$\left.\frac{d^2\widetilde{Q}_2}{d\sigma_n^2}\right|_{\sigma_n=\hat{\sigma}_{n,k+1}} = -8N^2\Big[\mathrm{Tr}\left(H^{\mathrm{H}} H\, F C_k\right) + \|\mathbf{y} - H\mathbf{m}_k\|^2\Big]^{-1} < 0, \tag{37}$$

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 initializing our estimate of σ_n to the median of the moduli of the finest-scale coefficients divided by $\sqrt{2\ln 2}$. This estimator for σ_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.

F. Summary of the Algorithm

We summarize our EM algorithm for image restoration above (we include references to equations earlier in the paper and, for notational convenience, we drop the subscript k).

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.

V. Computational Considerations

A. The Conditional Covariance Matrix C_k

In Section II-A, 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 (19). We recognize that calculating $\mathbf{m}_k$ is the same as solving the symmetric, positive-definite system $\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)\mathbf{m}_k = H^{\mathrm{H}}\mathbf{y}$, which we can do iteratively using a gradient-based method such as the conjugate gradients algorithm [27], [28]; this requires only multiplication by $\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)$, 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 $\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)$. The diagonal elements of $\hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}$ are trivial to calculate (as $\widehat{S}_k$ is diagonal), and we approximated each diagonal element of $H^{\mathrm{H}}H$ by the energy of the point-spread function 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 [27]:

$$\left\| \left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right) \mathbf{m}_k - H^{\mathrm{H}}\mathbf{y} \right\| \leq \tau \left(\left\| H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2} \right\|_F \|\mathbf{m}_k\| + \|H^{\mathrm{H}}\mathbf{y}\| \right), \quad (38)$$

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

- The Frobenius norm $\left\| H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2} \right\|_F$ can be expressed as $\sqrt{\mathrm{Tr}\left[\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^2\right]}$. Calculating it exactly is difficult, so instead we used the following unbiased stochastic estimator taken from [29], [30]: 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:

$$\mathrm{Tr}\left[\left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^2\right] = \mathrm{E}\left\{\mathbf{u}^{\mathrm{T}} \left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^2 \mathbf{u}\right\} \approx \mathbf{u}^{\mathrm{T}} \left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^2 \mathbf{u}. \quad (39)$$

- Selection of an appropriate stopping tolerance τ turns out to be quite crucial. A value of τ 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 τ that is too large results in underfiltered images being passed to the logarithmic denoising stage and introduces artifacts into the end result. In our experiments, we chose appropriate values of τ by trial-and-error.

The second occurrence of C_k is in (34) in which its diagonal elements $\left\{\sigma_{i,k}^2\right\}$ are needed to update the echogenicity estimate. Because we are unable, in practice, to explicitly specify C_k, we cannot access its diagonal elements $\left\{\sigma_{i,k}^2\right\}$ directly. We recognize, however, that each $\sigma_{i,k}^2$ is just the variance of $(\mathbf{x})_i$ given $\mathbf{y}$, $\widehat{S}_k$, and $\hat{\sigma}_{n,k}$. We can assume local ergodicity and estimate $\left\{\sigma_{i,k}^2\right\}$ by computing sample variances over local neighborhoods in some estimate of $\mathbf{x}$. As $\mathbf{m}_k$ is our best estimate of $\mathbf{x}$ given $\mathbf{y}$, $\widehat{S}_k$, and $\hat{\sigma}_{n,k}$, we expect a good estimate of $\left\{\sigma_{i,k}^2\right\}$ 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 (36) for $\hat{\sigma}_{n,k}$, in the term $\mathrm{Tr}\left(H^{\mathrm{H}}H\, C_k\right)$. We can use an approximation similar to (39) and write:

$$\begin{aligned}\mathrm{Tr}\left(H^{\mathrm{H}}H\, C_k\right) &= \mathrm{E}\left[\mathbf{u}^{\mathrm{T}} H^{\mathrm{H}}H\, C_k \mathbf{u}\right] \\ &\approx \mathbf{u}^{\mathrm{T}} H^{\mathrm{H}}H\, C_k \mathbf{u} \\ &= 2\hat{\sigma}_{n,k}^2 \mathbf{u}^{\mathrm{T}} H^{\mathrm{H}}H \left(H^{\mathrm{H}}H + \hat{\sigma}_{n,k}^2 \widehat{S}_k^{-2}\right)^{-1} \mathbf{u}.\end{aligned} \quad (40)$$

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

B. 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 (33) and (34) beyond stating that it is to be specified in terms of the wavelet coefficients of the echogenicity image. Therefore, we are free to use any wavelet transform and any shrinkage rule we choose.

We recommend using the dual-tree complex wavelet transform (DTCWT) [31]–[33] over the conventional DWT because it has better properties than the DWT for image processing that comes 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 [31].

Unlike the DWT, the DTCWT is redundant and, hence, not orthonormal. However, it is 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 that can fluctuate significantly with spatial shifts.
- Directional selectivity: In two dimensions or higher, the DTCWT is able to localize 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 [25], [26] 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).

VI. Experimental Results

In this section, we present a number of results on synthetic and *in vitro* images that compare the performance of our algorithm with zero-order Tikhonov regularization, l^1-norm (Laplacian) regularization, and ForWaRD. To perform the necessary function minimizations, we used the conjugate gradients algorithm with diagonal preconditioning. Expressions for the gradient vector and Hessian matrices used in the conjugate gradients algorithm are listed in Appendix C.

In each of our experiments, we approximated a linear shift-variant blurring operator by partitioning image space axially into a number of regions of equal size and simulating the response to a point scatterer in the center of each region in Field II [1]. We used the parameters for a linear 5–10 MHz array with 127 elements (spanning a lateral length of 40 mm), an active aperture of 32 elements, and a single coincident lateral focus on both transmission and reception. The lateral focal length varied, depending on the experiment, but elevational focusing was achieved with a fixed acoustic lens of focal length of 23 mm. The point-spread functions were scaled to give the point-spread function closest to the lateral focus unit energy. We estimated the additive noise variance using the wavelet estimator described in Section IV-E. All the methods we tested require a value for the global image variance σ_x^2 that we approximated as the variance of the blurred image less the estimated noise variance. With ForWaRD, we used an underregularization parameter of $\alpha = 0.2$ as prescribed in [19], and we applied Sendur and Selesnick's bivariate [25], [26] shrinkage rule for the subsequent wavelet denoising. The bivariate shrinkage rule requires the variance of each wavelet coefficient to be specified, and we approximated these variances according to [25], [26]. We also used the bivariate shrinkage rule in the M-step of our proposed algorithm; and we initialized the variances of the wavelet coefficients from the zero-order Tikhonov-regularized image produced in the first E-step.

A. 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 equalized and scaled to give a signal-to-noise ratio in the logarithmic domain of 20 dB. We used eight point-spread functions and blurred the tissue reflectivities in the manner described in Section III and we added complex-valued, white Gaussian noise to give the resulting images a blurred signal-to-noise ratio (BSNR) of 20 dB.

We quantified the performance of each algorithm according to its result's improvement in signal-to-noise ratio (ISNR) in decibels, calculated as:

$$\text{ISNR} = 20\log_{10}\left(\frac{\|\mathbf{y}-\mathbf{x}\|}{\|\hat{\mathbf{x}}-\mathbf{x}\|}\right), \tag{41}$$

where $\mathbf{x}$ is the vectorized tissue reflectivity, $\mathbf{y}$ is the noise and blurred image, and $\hat{\mathbf{x}}$ is the vectorized 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 (The MathWorks, Inc., Natick, MA) on a personal computer with a 3.2 GHz processor and 1 GB of memory.

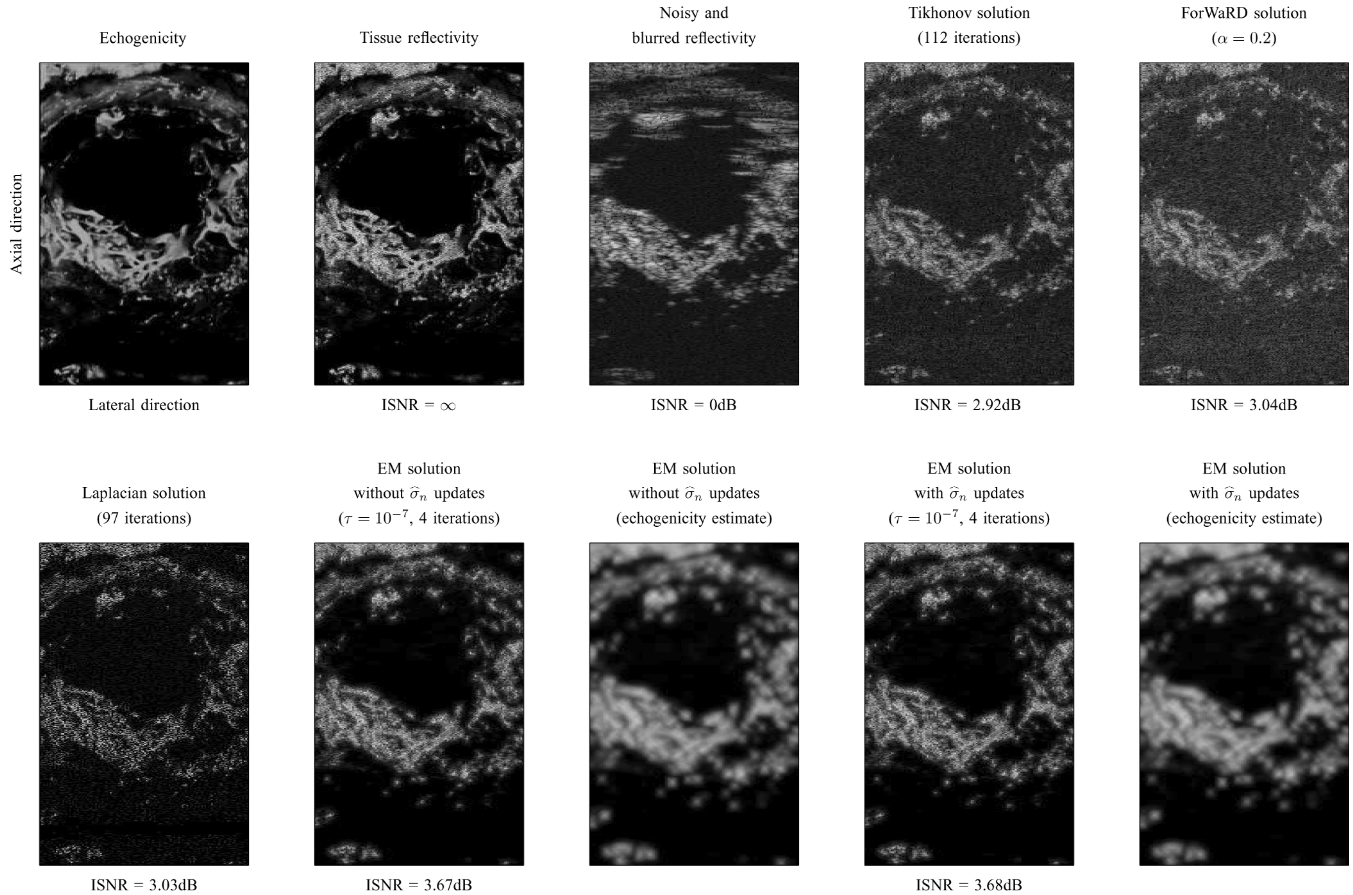
1. Heart Image: Our first set of simulations was conducted on a photograph of a cross section through a human heart. The results of the various restoration methods are shown in Fig. 2(a), and a plot of the evolution of ISNR against execution time for the zero-order Tikhonov, Laplacian, and EM restorations is shown in Fig. 2(b) (we have omitted ForWaRD from this plot because its ISNR can be sensibly computed only 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 0.63 dB 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).

2. Kidney Image: Our second set of simulations was conducted on an artist's illustration of a cross section through a human 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 0.61 dB 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.

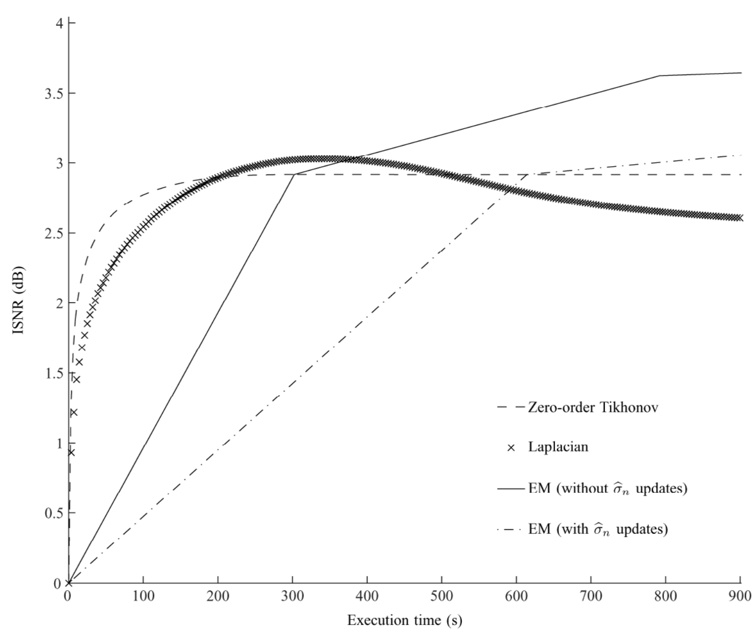
B. In Vitro Results

We acquired an image of a phantom containing spherical inclusions of different echogenicities embedded in a background of dense scatterers with the linear array described at the start of this section and a Dynamic Imaging (Livingston, Scotland, UK) Diasus ultrasound machine. Standard delay-and-sum beamforming was applied and the beamformed traces were sampled at a rate of 66.6 MHz. Each sampled trace was then demodulated to baseband, low-pass filtered, and downsampled by a factor of nine.

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 regularization, ForWaRD, and Laplacian regularization. Our approximation of the shift-variant blurring operator with locally shift-invariant point-spread functions 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 in which lateral focusing 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.

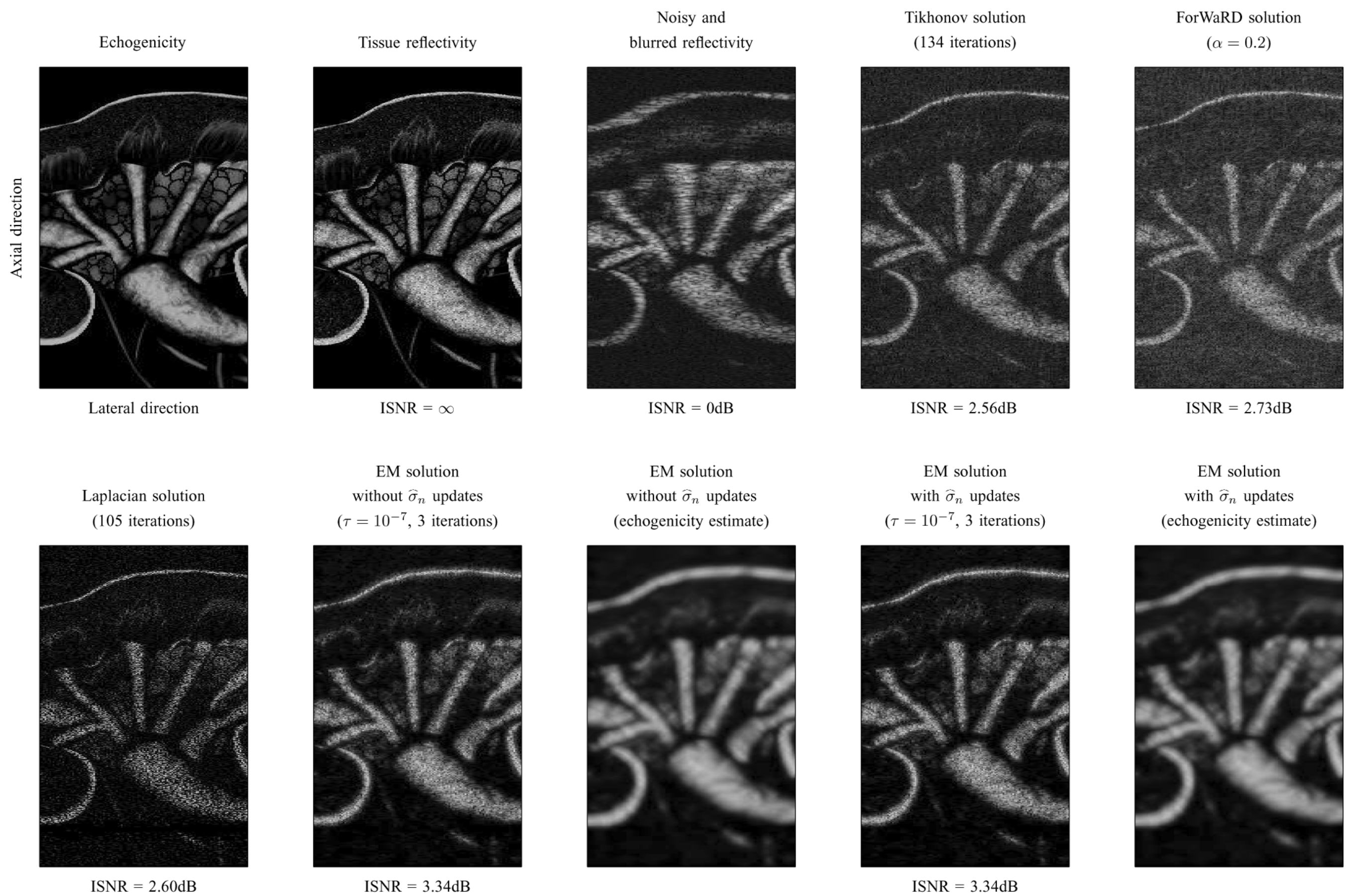


(a)

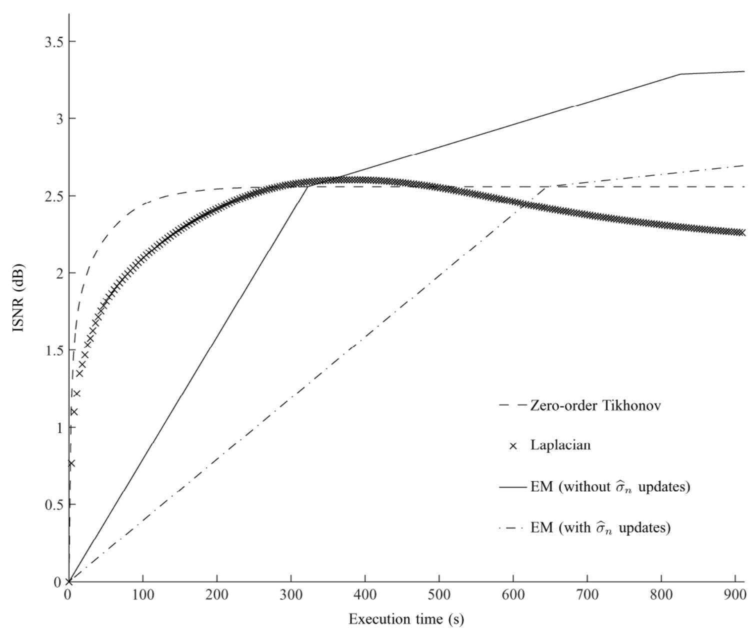


(b)

Fig. 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 40 dB. The lateral focus is approximately at the center of the image. The echogenicity image was obtained with permission from www.umdnj.edu/pathnweb/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 3.2 GHz processor and 1 GB of memory.



(a)



(b)

Fig. 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 40 dB. The lateral focus is approximately in the center of the image. The echogenicity image was obtained with permission from 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 3.2 GHz processor and 1 GB of memory.

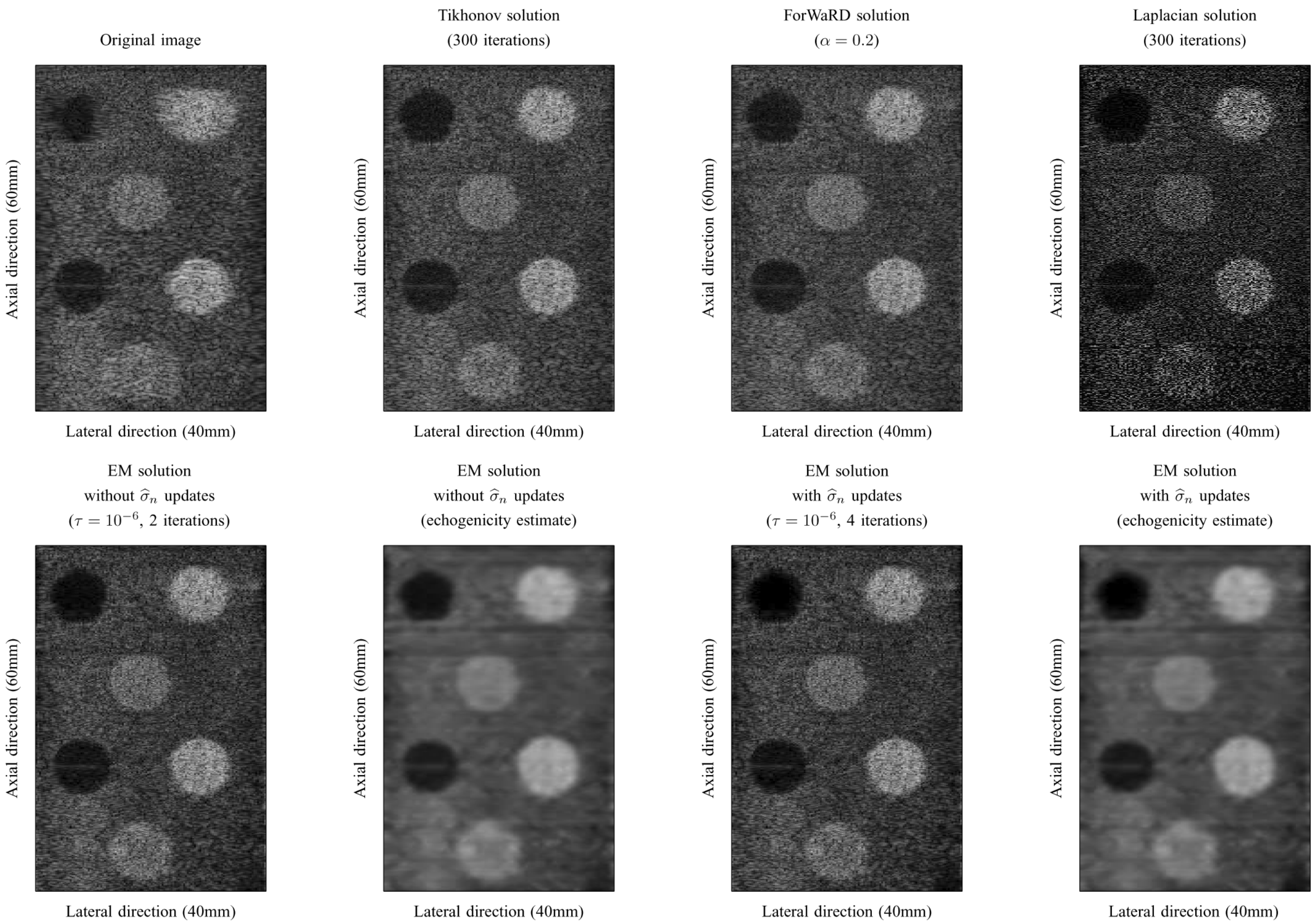


Fig. 4. *In vitro* results for an image of a phantom containing spherical inclusions. The lateral focus is 26 mm below the top of the image. The grey-scale levels represent logarithmically compressed amplitudes and span a dynamic range of 40 dB.

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 decibels) are listed in Table II. To calculate each contrast, we computed the energy of the pixels in a rectangular area in each of two adjacent regions and took the ratio of the energies. 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.

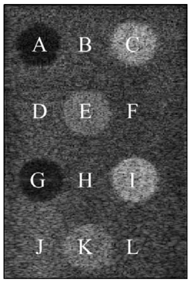
VII. Conclusions

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, although 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. Therefore, we proposed modeling 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 modeled as a natural image that 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

TABLE II

LOCAL CONTRASTS IN THE ORIGINAL AND RESTORED IMAGES OF THE PHANTOM CONTAINING SPHERICAL INCLUSIONS.[1]



	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

[1]The image on the left labels 12 regions in the phantom and the table on the right contains contrast values (in decibels) between selected pairs of adjacent 6.25×6.25 mm regions. All contrast values were calculated from the restored images after artifact correction.

(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 regularization 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 in which 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 it 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 optimizations as l^1- and l^2-norm regularization. Hence, the benefits introduced by our algorithm come with only a small computational penalty.

We also have included in this paper closed-form expressions for the statistics of the logarithm of a Rayleigh-distributed random variable [(32) and Appendix A] and a robust wavelet-based estimator for the variance of complex-valued additive white Gaussian noise in a realistic signal (Section IV-E and Appendix B). We have been unable to find these expressions in the technical literature, so we have presented them here for the sake of completeness.

APPENDIX A
DERIVATION OF THE STATISTICS OF LOGARITHMIC NOISE

In this appendix, we derive the probability density function and the statistics quoted in (32) for the logarithmic noise term $\{\widetilde{w}_i\}$ in (31). We first state the following two useful equations:

$$\int_0^\infty e^{-u} \ln u \, du = -\gamma, \tag{42}$$

$$\int_0^\infty e^{-u} (\ln u)^2 \, du = \gamma^2 + \frac{\pi^2}{6}, \tag{43}$$

where γ is the Euler-Mascheroni constant. Expressions (42) and (43) are, respectively, special cases of (4.331-1) and (4.335-1) in [34].

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

Proof: Define $W = |Z|$. Marginalizing the probability density function of Z over $\arg(Z)$ (see, for example, [35]), 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 $\widetilde{W} = \ln W$. The probability density function of $\widetilde{W}$ therefore is:

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

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

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

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

The first integral evaluates to 1 and, from (42), the second integral evaluates to $-\gamma$. Hence, $\mathrm{E}(\widetilde{W}) = \frac{1}{2}(\ln 2 - \gamma)$.

Proposition 3: Given a complex-valued Gaussian random variable Z with mean $\mathrm{E}(Z) = 0$ and variance $\mathrm{E}\left(|Z|^2\right) = 2$, and defining $\widetilde{W} = \ln|Z|$, the variance of $\widetilde{W}$ is given by $\mathrm{Var}\left(\widetilde{W}\right) = \frac{\pi^2}{24}$.

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

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

As before, the first integral evaluates to 1, and, from (42) and (43), the second and third integrals evaluate to $-\gamma$ and $\gamma^2 + \frac{\pi^2}{6}$ respectively. So $\mathrm{E}\left(\widetilde{W}^2\right) = \frac{1}{4}\left[(\ln 2)^2 - 2\gamma\ln 2 + \gamma^2 + \frac{\pi^2}{6}\right]$, and:

$$\begin{aligned} \mathrm{Var}\left(\widetilde{W}^2\right) &= \mathrm{E}\left(\widetilde{W}^2\right) - \left[\mathrm{E}\left(\widetilde{W}\right)\right]^2 \\ &= \frac{1}{4}\left[(\ln 2)^2 - 2\gamma\ln 2 + \gamma^2 + \frac{\pi^2}{6}\right] \\ &\quad - \frac{1}{4}\left[(\ln 2)^2 - 2\gamma\ln 2 + \gamma^2\right] = \frac{\pi^2}{24}. \end{aligned}$$

Appendix 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 IV-E. Our derivation is analogous to the derivation of a similar robust estimator for the variance of real-valued AWGN presented in [9].

Proposition 4: Given a complex-valued Gaussian random variable Z with mean $\mathrm{E}(Z) = 0$ and variance $\mathrm{E}\left(|Z|^2\right) = 2\sigma_n^2$, and defining $W = |Z|$, the median of W is $\sigma_n\sqrt{2\ln 2}$.

Proof: The calculation is fairly straightforward. As in Appendix A, if we marginalize 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:

$$\begin{aligned} & \int_{-\infty}^{w_m} p_W(w) dw = \frac{1}{2} \\ \Rightarrow & \int_0^{w_m} \frac{w}{\sigma_n^2}\exp\left(-\frac{w^2}{2\sigma_n^2}\right) dw = \frac{1}{2} \\ \Rightarrow & \qquad w_m = \sigma_n\sqrt{2\ln 2}. \end{aligned}$$

Given a realistic (i.e., band-limited) signal contaminated by complex-valued, additive, white Gaussian noise, we expect its finest-scale DWT coefficients to be dominated by the noise, and because the DWT is orthonormal, the finest-scale coefficients are also likely to be distributed as complex-valued Gaussian random variables with the same variance as the noise. The sample median of the moduli of these coefficients generally is immune to any potential outliers and is approximately equal to the population median $\sigma_n\sqrt{2\ln 2}$, so a robust estimate of σ_n is this sample median divided by $\sqrt{2\ln 2}$.

Appendix 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 minimization requires knowledge of the gradient vector and the Hessian matrix of the cost function to be minimized. The cost functions in (8) and (10) 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 [20], [36], and [37].

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 [36], 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:

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

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

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 $\phi : \mathbf{x} \in \mathbb{C}^N \to \mathbb{R}$ is then defined to be:

$$\nabla\phi \equiv 2\frac{\partial \phi}{\partial \mathbf{x}^*},$$

and is normal to the surface of the cost function [36]. We similarly define the Hessian matrix $\mathcal{H}_\phi$ of ϕ to be:

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

The cost functions we encounter in (8) and (10) are of the form:

$$\phi(\mathbf{x}) = \frac{1}{2\sigma_n^2}\|\mathbf{y} - H\mathbf{x}\|^2 + r(\mathbf{x}) = \frac{1}{2\sigma_n^2}\left(\mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}H\mathbf{x} - \mathbf{x}^{\mathrm{H}}H^{\mathrm{H}}\mathbf{y} + \mathbf{x}^{\mathrm{H}}H^{\mathrm{H}}H\mathbf{x}\right) + r(\mathbf{x}),$$

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

$$\nabla\phi = \sigma_n^{-2}\left(H^{\mathrm{H}}H\mathbf{x} - H^{\mathrm{H}}\mathbf{y}\right) + \nabla r,$$
$$\mathcal{H}\phi = \sigma_n^{-2}H^{\mathrm{H}}H\mathbf{x} + \mathcal{H}r.$$

In the case of l^2-norm regularization (8), $r(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{H}}C_2^{-1}\mathbf{x}$ and $\nabla r = C_x^{-1}\mathbf{x}$ and $\mathcal{H}r = C_x^{-1}$.

In the case of l^1-norm regularization (10), $r(\mathbf{x}) = \sqrt{3}\sigma_x^{-1}\sum_{i=1}^N |x_i| = \sqrt{3}\sigma_x^{-1}\sum_{i=1}^N \sqrt{x_i x_i^*}$, and the gradient vector and Hessian matrix of $r(\mathbf{x})$ are given componentwise by:

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

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

APPENDIX 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 nonexact 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, it 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 the first step 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 equalize 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).

ACKNOWLEDGMENTS

James Ng is funded by an External Research Studentship from Trinity College, Cambridge, and an LB Wood Travelling Scholarship.

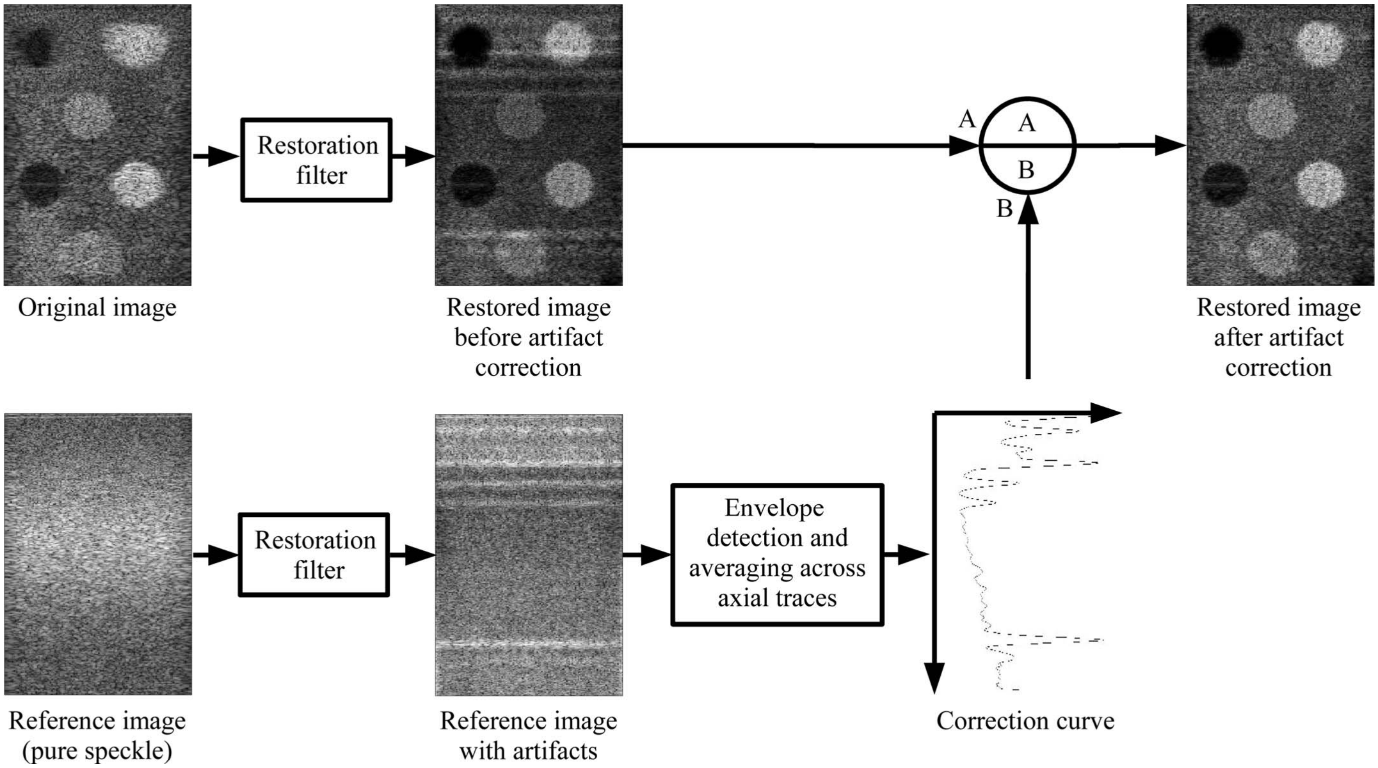


Fig. 5. Correction of artifacts introduced by errors in the approximation to the blurring operator and exacerbated by image restoration. Applying the restoration filter to a reference image of uniformly echogenic speckle produces an image that exhibits artifacts from the blurring operator approximation. 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 40 dB.

Graham Treece is funded by a Royal Academy of Engineering/EPSRC Research Fellowship.

References

[1] J. A. Jensen, "Speed-accuracy trade-offs in computing spatial impulse responses for simulating medical ultrasound imaging," *J. Comput. Acoust.*, vol. 9, no. 3, pp. 731–744, 2001.

[2] O. Michailovich and D. Adam, "A novel approach to the 2-D blind deconvolution problem in medical ultrasound," *IEEE Trans. Med. Imag.*, vol. 24, pp. 86–104, Jan. 2005.

[3] 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, pp. 1286–1295, Oct. 2003.

[4] T. Taxt and J. Strand, "Two-dimensional noise-robust blind deconvolution of ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, pp. 861–866, July 2001.

[5] T. Taxt, "Three-dimensional blind deconvolution of ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, pp. 867–871, July 2001.

[6] Y. Li, "Phase aberration correction using near-field signal redundancy—Part I: Principles," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 44, pp. 355–371, Mar. 1997.

[7] 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, pp. 372–379, Mar. 1997.

[8] L. Nock, G. E. Trahey, and S. W. Smith, "Phase aberration correction in medical ultrasound using speckle brightness as a quality factor," *J. Acoust. Soc. Amer.*, vol. 85, pp. 1819–1833, May 1989.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[10] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vision*, vol. 18, pp. 17–33, 2003.

[11] 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, pp. 297–311, Mar. 2006.

[12] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Trans. Med. Imag.*, vol. 25, pp. 229–240, Feb. 2006.

[13] O. V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 53, pp. 64–78, Jan. 2006.

[14] 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, pp. 1653–1669, Oct. 2005.

[15] 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, pp. 743–754, June 2005.

[16] 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, pp. 121–130, Jan. 2001.

[17] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 12, pp. 906–916, Aug. 2003.

[18] R. Neelamani, H. Choi, and R. Baraniuk, "ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Trans. Signal Processing*, vol. 52, pp. 418–433, Feb. 2004.

[19] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-domain regularized deconvolution for ill-conditioned systems," in *IEEE Int. Conf. Image Processing*, Oct. 1999, pp. 204–208.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory, ser. Signal Processing.* vol. 1, Englewood Cliffs, NJ: Prentice-Hall, 1993, pp. 500–524.

[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 VII.* Int. Soc. Opt. Eng., 1997, pp. 388–399.

[22] J. G. Nagy and D. P. O'Leary, "Restoring images degraded by spatially-variant blur," *SIAM J. Sci. Comput.*, vol. 19, pp. 1063–1082, 1998.

[23] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging.* Bristol: Institute of Physics Publishing, 1998.

[24] F. Dellaert, "The expectation maximization algorithm," Atlanta, GA: Georgia Institute of Technology, Technical Report GIT-GVU-02-20, Feb. 2002.

[25] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Lett.*, vol. 9, pp. 438–441, Dec. 2002.

[26] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Processing*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.

[27] 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 ed. Philadelphia, PA: SIAM, 1994, pp. 57–63.

[28] 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: Cambridge Univ. Press, 2002, pp. 74–92.

[29] K. M. Perry and S. J. Reeves, "A practical stopping rule for iterative signal restoration," *IEEE Trans. Signal Processing*, vol. 42, pp. 1829–1833, July 1994.

[30] M. F. Hutchinson, "A stochastic estimator for the trace of the influence matrix for Laplacian smoothing splines," *Commu. Statistics—Simulation Comput.*, vol. 19, no. 2, pp. 433–450, 1990.

[31] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Mag.*, vol. 22, pp. 123–151, Nov. 2005.

[32] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Comput. Harmonic Anal.*, vol. 10, pp. 234–253, 2001.

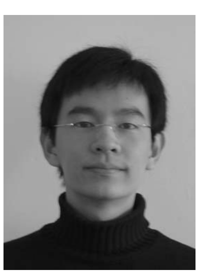
[33] N. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. R. Soc. Lond. A*, vol. 357, pp. 2543–2560, 1999.

[34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products.* 5th ed. A. Jeffrey, Ed. London: Academic Press, 1994.

[35] 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, pp. 549–563, Mar. 2006.

[36] S. Haykin, *Adaptive Filter Theory.* 3rd ed. ser. Information and System Sciences Englewood Cliffs, NJ: Prentice-Hall, 1996, pp. 890–894.

[37] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc. H (Microwaves, Optics Antennas)*, vol. 130, no. 1, pp. 11–16, Feb. 1983.



James Ng was born in Kuala Lumpur, Malaysia, in 1981. From 1999 to 2002 he studied at the University of Auckland, Auckland, New Zealand, obtaining the B.E. (Hons) degree in computer systems engineering in May 2003. He currently holds an External Research Studentship from Trinity College, University of Cambridge, Cambridge, UK, and an LB Wood Travelling Scholarship. Since October 2003, he has been conducting research towards the Ph.D. degree at the University of Cambridge's Department of Engineering. His research interests are in image processing, complex wavelets, and inverse problems. He is at present working on Bayesian methods for the restoration of medical ultrasound images.

Richard Prager received the B.A. degree in electrical sciences at the University of Cambridge (UK) in 1984 and the Ph.D. degree in Information Engineering in 1987. He was elected to a Fellowship at Queens' College Cambridge in 1988. In the Department of Engineering of the University of Cambridge he was appointed Senior Assistant in Research in 1989, Lecturer in 1992, Senior Lecturer in 2000 and Reader in Information Engineering in 2002. From 1996 to 1999 he held the post of Director of Undergraduate Education and in 2001 was awarded the Pilkington Prize for excellence in teaching by the Trustees of the Cambridge Foundation. His research interests are in medical imaging and include acquisition and use of radio-frequency three-dimensional ultrasound. He is one of the principal architects of the both the Stradx and Stradwin freehand 3-D ultrasound systems. He is a Chartered Engineer (CEng) through membership of the Institute of Electrical Engineers.

Nick Kingsbury (M'87) received the honours degree in 1970 and the Ph.D. degree in 1974, both in electrical engineering, from the University of Cambridge. He is a member of the IEEE.

From 1973 to 1983 he was a Design Engineer and subsequently a Group Leader with Marconi Space and Defence Systems, Portsmouth, England, specializing in digital signal processing and coding, as applied to speech coders, spread spectrum satcomms, and advanced radio systems. Since 1983 he has been a Lecturer in Communications Systems and Image Processing at the University of Cambridge and a Fellow of Trinity College, Cambridge. He was appointed to a Readership in Signal Processing at Cambridge in 2000.

His current research interests include image compression, error-robust source coding techniques, and image analysis and enhancement techniques, particularly those based on wavelet decompositions. He has developed the dual-tree complex wavelet transform and is especially interested in the application of complex wavelets to the analysis of images and 3-D datasets.

Graham Treece was awarded a first class B.A. degree in Engineering, with distinction, from the University of Cambridge, in 1993. From 1994 to 1995 he worked for Noise Cancellation Technologies Ltd., Cambridge, UK, designing active noise cancellation prototypes for cars and trucks. After this, he was employed by Plextek Ltd., Great Chesterford, Essex, UK, a leading independent design consultancy specialising in microelectronics and communications, as a real-time embedded software engineer. In 1997 he returned to the University of Cambridge, and received his Ph.D. on volume measurement and surface visualisation in 3-D ultrasound in June 2001. Since then he has continued to work in this area, initially as a Research Associate. In October 2003 he was awarded a Research Fellowship from the Royal Academy of Engineering and the Engineering and Physical Sciences Research Council in the UK. He is also a Fellow of Queens' College, Cambridge.

His current research interests are in visualisation, system design and the reduction of artifacts in 3-D ultrasound. He is one of the authors of the Stradx 3-D ultrasound system, widely recognized as state-of-the-art. He also has research interests in medical imaging and computer graphics, particularly surface visualisation.

Andrew Gee obtained a B.A. in Electrical and Information Science at the University of Cambridge in 1990. He was awarded the Ph.D. degree in 1993 from the same university for his thesis on the use of feedback neural networks for combinatorial optimization. In 1993, he was appointed to a Research Fellowship at Queens' College, Cambridge, where he started to work on computer vision. In October 1995 he became an Assistant Lecturer in Information Engineering and an Official Fellow of Queens' College, where he is now Director of Studies in Engineering. He was promoted to the post of Lecturer in January 1998 and Senior Lecturer in October 2001. He teaches undergraduate courses in all aspects of computing: in 1998 he was awarded a Pilkington Prize by the Trustees of the Cambridge Foundation for excellence in teaching at the University of Cambridge. He has worked in a number of research areas, including medical imaging, computer vision, document image processing, and neural networks. Currently, he is working on aspects of three-dimensional ultrasound and related topics in computer graphics. He has authored and co-authored around 100 papers.