



PAPER

The convolutional interpretation of registration-based plane wave steered pulse-echo local sound speed estimators

Anthony S Podkova¹  and Michael L Oelze

Department of Electrical and Computer Engineering, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, United States of America

¹ Author to whom any correspondence should be addressed.E-mail: tpodkow2@illinois.edu**Keywords:** sound speed, ultrasound computed tomography, inverse problemsRECEIVED
3 April 2019REVISED
13 November 2019ACCEPTED FOR PUBLICATION
9 December 2019PUBLISHED
16 January 2020**Abstract**

Pulse-echo reconstruction of sound speed has long been considered a difficult problem within the domain of quantitative biomedical ultrasound. However, recent results (Jaeger 2015 *Ultrasound Med. Biol.* **41** 235–50; Jaeger and Frenz 2015 *Ultrasonics* **62** 299–304; Jaeger *et al* 2015 *Phys. Med. Biol.* **60** 4497–515) have demonstrated that pulse-echo reconstructions of sound speed are achievable by exploiting correlations in post-beamformed data from steered, plane-wave excitations in the presence of diffuse scatterers. Despite these recent advances, a coherent theoretical imaging framework for describing the approach and results is lacking in the literature. In this work, the problem of sound speed reconstruction using steered, plane-wave excitations is reformulated as a truncated convolutional problem, and the theoretical implications of this reformulation are explored. Additionally, a matrix-free algorithm is proposed that leverages the computational and storage advantages of the fast Fourier transform (FFT) while simultaneously avoiding FFT wraparound artifacts. In particular, the storage constraints of the approach are reduced down from $\mathcal{O}(M^2N^2)$ to $\mathcal{O}(MN)$ over full matrix reconstruction, making this approach a better candidate for large reconstructions on clinical machines. This algorithm was then tested in the open source simulation package *k*-Wave to assess its robustness to modeling error and resolution reduction was demonstrated under full-wave propagation conditions relative to ideal straight-ray simulations. The method was also validated in a phantom experiment.

1. Introduction

Ultrasonography is one of the most widely used modalities in the field of biomedical imaging due to its low cost, safety, portability, and real-time capabilities. However, conventional ultrasonography is plagued by a number of practical limitations. For example, a couple of drawbacks associated with conventional brightness mode (B-mode) imaging are its mainly qualitative nature and the presence of speckle artifacts, which limits the capability to observe contrast. Furthermore, due to inherent system and operator dependencies for clinical sonography, it is often difficult to compare B-mode images acquired with different systems, transducers, and setting configurations. For this reason, the diagnostic capabilities of B-mode ultrasound can be highly subjective.

In order to address the qualitative nature of ultrasonic imaging, a number of quantitative acoustic imaging modes have been proposed. These modes attempt to address the inherent system and operator dependency issues by directly estimating physically relevant acoustic properties such as attenuation, backscatter coefficient/reflection, and sound speed, in a way that at least partially accounts for the limitations of the measurement apparatus (Norton and Linzer 1981, Mamou and Oelze 2013, Oelze and Mamou 2016). For the purposes of this work, the scope will be restricted to the quantitative imaging of sound speed, an acoustic parameter that has been shown to be effective at differentiating between benign and malignant tissue states in both the liver and breast, among others (Bamber and Hill 1981, Bamber *et al* 1981, Sehgal *et al* 1986, Duric *et al* 2007, Johnson *et al* 2007).

From a comparative methodological standpoint, the most definitive methods for imaging sound speed are those of traditional ultrasonic computed tomography (USCT) (Greenleaf *et al* 1975, Glover and Sharp 1977,

Duric *et al* 2007, Wiskin *et al* 2007, Lavarello and Oelze 2008, 2009, Hormati *et al* 2010, Li *et al* 2010, Huthwaite and Simonetti 2011). The most salient feature of these methods is the acquisition of the forward propagating wave through a transmission configuration where the source and receiver are located on opposite sides of the sample. This can be accomplished using two mechanically translated single-element transducers (Greenleaf *et al* 1975, Glover and Sharp 1977), two diametrically opposed linear arrays (Johnson *et al* 2007), or a single ring array (Carson *et al* 1981, Duric *et al* 2007, Hormati *et al* 2010, Li *et al* 2010, Huthwaite and Simonetti 2011). The main advantage of such configurations is the preservation of baseband information, because transmission measurements sample the low frequencies of the Ewald sphere (Blahut 2004, Devaney 2012), which also have the strongest signal-to-noise ratio (SNR) under Born scattering. However, such configurations are unwieldy in practice, require additional custom hardware, and are practically limited in scope to easily externalized soft tissues such as the female breast.

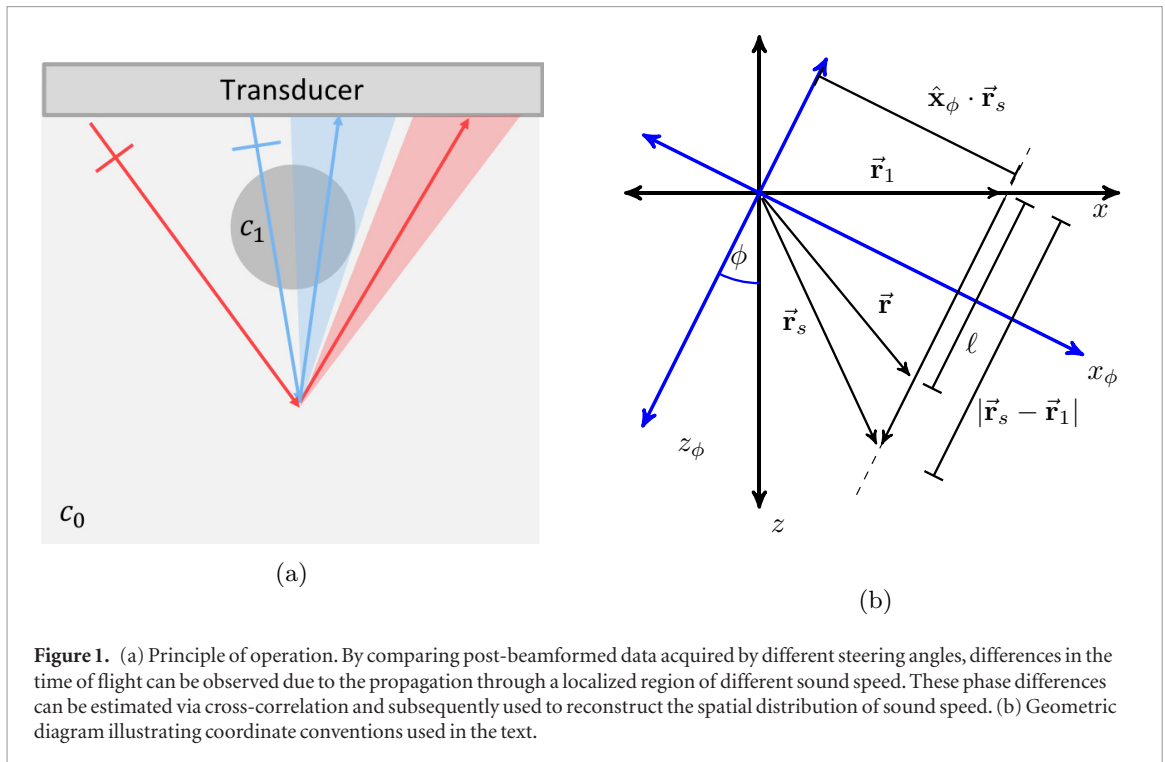
For this reason, it is desirable to develop quantitative methods that can be deployed in a pulse-echo environment, which is achievable using conventional linear arrays. The main drawback of the pulse-echo constraint is a reduction in data availability. In a pulse-echo setting, the information reduction manifests as a bandpass filtering effect, due to a limited sampling of the Ewald sphere (Norton and Linzer 1981, Lu *et al* 1994, Lu 1997, Walker and Trahey 1998). It is for this reason that pulse-echo sound speed imaging in the general case is considered difficult because the baseband components are not preserved in the data. This limitation also impacts the quality of B-mode images, because interactions between sub-resolution scatterers ultimately manifest as speckle after baseband demodulation.

Despite these inherent difficulties, there have been several attempts at echo-mode sound speed imaging. One class of methods for pulse-echo sound speed imaging is based around the beam tracking methodology (Ophir 1986, Kondo *et al* 1990, Céspedes *et al* 1992), which exploits registered backscattered data from two spatially separated transducers to reconstruct local sound speed. These methods rely on the presence of diffuse scatterers throughout the medium to make inferences about local sound speed. In another approach, backscattered data are registered using parabolic fits of prebeamformed radio frequency (RF) echoes to generate sound speed reconstructions under both focused transmit and single-element excitations (Byram *et al* 2012, Jakovljevic *et al* 2018). The applicability of these methods is likely to be limited by motion artifacts due to the number of excitations required to generate a full dataset.

More recently, the beam-tracking methodology was generalized to post-beamformed steered plane-wave data acquired by linear array transducers (Jaeger and Frenz 2015, Jaeger *et al* 2015a, 2015b, Sanabria *et al* 2018). Plane-wave steering forms the basis for the current standard of high frame rate ultrasound, because it is able to rapidly sample Fourier space in a way that minimizes the total number of excitations while preserving spatial resolution (Lu *et al* 1994, Lu 1997, Montaldo *et al* 2009). It is precisely this property that makes plane-wave imaging advantageous for Doppler ultrasound applications, and suggests potential robustness to motion artifacts compared to translated transmit aperture methods (Byram *et al* 2012, Jakovljevic *et al* 2018).

However, theoretical characterizations of pulse-echo sound speed reconstruction using steered plane waves in a broader imaging context are lacking. A first order model of the Fourier-space coverage was presented in Jaeger *et al* (2015a), and was used to derive an FFT-based reconstruction algorithm. Results from this paper have been difficult to reproduce completely. This is unfortunate, as the core idea of utilizing local phase estimates to extract information about sound speed by varying transmit direction is worthwhile. As an alternative, a spatial domain approach of the problem was developed in Sanabria *et al* (2018), which enabled the incorporation of spatial domain total variation regularization into the method, as well as the option to incorporate manually segmented prior information into the reconstruction. However, the approach is nonlinear and biased towards piecewise constant reconstructions. Furthermore, only two results were presented under full-wave propagation conditions, with the vast majority of the results being on straight ray simulated data which would exhibit a quality bias consistent with inverse crime. Of the two realistic simulations, the only validated full-wave result was a simulation developed in house, which could have been unintentionally biased in favor of their algorithm.

The main contributions of this work are as follows. First, the problem of registration-based plane-wave sound speed imaging is reformulated in a convolutional framework, in order to aid future researchers interested in implementing such methods. Second, the Fourier-domain characterization is directly related to the Fourier projection-slice theorem and the traditional backscattered Fourier coverage as one might expect from a diffraction tomographic approach or traditional B-mode ultrasound. Third, a matrix-free algorithmic approach based off of the convolutional reformulation is proposed which avoids the FFT wraparound artifacts while simultaneously leveraging the computational and storage advantages for the FFT. The benefits of matrix-free reconstructions are well-known in the literature, and have been concisely summarized in Diamond and Boyd (2016). This algorithm reduces storage constraints from $\mathcal{O}((MN)^2)$ to $\mathcal{O}(MN)$, which enables larger reconstructions to be performed on clinical machines. This method was tested in the open source simulation package *k-Wave* (Treeby *et al* 2012) to assess its robustness to modeling error and the resultant reconstructions exhibited significant resolution degradation relative to straight-ray modeling conditions. Similar results were achieved experimentally in a



phantom study. Finally, the main advantages and limitations of the method are identified, and areas of potential future research are explored.

2. Methods

2.1. Convolutional formulation of the straight ray model

The principle of operation of the methodology is illustrated in figure 1(a). Consider the propagation of two steered plane waves in a heterogeneous medium. A first wave (red) is transmitted at one angle and the scattered field recorded at the transducer array and the second wave (blue) is transmitted at a different angle and received by the transducer array. The heterogeneous medium is assumed to be comprised of both diffusely distributed small scale weak Born scatterers ($\ll \lambda$) as well as large scale aberrating inclusions ($\gg \lambda$, Rytov-like scatterers (Devaney 2012, Chew 1995)), which serve to distort the phase of the forward propagating waves. The received echoes are beamformed using traditional delay and sum methodology (Montaldo *et al* 2009) and the local phase differences are compared using a cross-correlation approach similar to that used by Jaeger *et al* (2015a). In the case of the blue wavefront, the backscattered signal will be phase-shifted due to the inclusion, whereas the data from the red wavefront will not. Thus, by modulating the transmit steering angle of the aperture and comparing the local phase differences of the received echoes via cross-correlation, forward scatter information is gained about the time-of-flight differences between the different paths. We desire to form a model that appropriately characterizes lags estimated by cross-correlation of two post-beamformed frames with transmit steering angles α and β . The relative lag estimated at any given point in the region of interest will be the difference between total deviations in time of flight for each frame:

$$\Delta\tau(\vec{r}_s; \alpha, \beta) = \tau_T(\vec{r}_s; \alpha) - \tau_T(\vec{r}_s; \beta) \quad (1)$$

where \vec{r}_s is the location of the scatterer.

The total time of flight from a transmitting aperture to a scatterer and back can be further decomposed into propagation times corresponding to the transmit and receive paths. It should be noted that the influence of the receive path has been largely neglected in the published literature (Jaeger and Frenz 2015, Jaeger *et al* 2015a, 2015b, Sanabria *et al* 2018). For simplicity of analysis, we will restrict our analysis to the case where the steering angle of the receive aperture is mirrored with respect to the transmit (i.e. $\phi_{Rx} = -\phi_{Tx}$), as this keeps the 2D spectrum approximately invariant for correlation processing (Walker and Trahey 1998) which is examined in further detail in section 2.4. Thus, the relative lag estimated from cross-correlation is a superposition of four terms:

$$\Delta\tau(\vec{r}_s; \alpha, \beta) = \underbrace{[\tau(\vec{r}_s; \alpha) + \tau(\vec{r}_s; -\alpha)]}_{\tau_T(\vec{r}_s; \alpha)} - \underbrace{[\tau(\vec{r}_s; \beta) + \tau(\vec{r}_s; -\beta)]}_{\tau_T(\vec{r}_s; \beta)}. \quad (2)$$

Each path can be written in the form of a line integral of the perturbation of inverse sound speed, or slowness

$$\tau(\vec{r}_s; \phi) = \int_{\vec{r}_1(\phi)}^{\vec{r}_s} d\ell \sigma(\vec{r}), \quad (3)$$

$$= \int_0^{|\vec{r}_s - \vec{r}_1|} d\ell \sigma(\vec{r}_1 + \ell \hat{z}_\phi) \quad (4)$$

where $\sigma(\vec{r}) = 1/c(\vec{r}) - 1/c_{bf}$ is the spatially varying slowness perturbation, $c(\vec{r})$ is the spatially varying sound speed, c_{bf} is the *a priori* sound speed used for beamformation, $d\ell$ is the differential path length, ϕ is the steering angle of the aperture and $\vec{r}_1(\phi)$ is the corresponding aperture centroid as appropriate. This integral can be reexpressed in a convolutional form

$$\tau(\vec{r}_s; \phi) = \sigma(\vec{r}_s) *_{\vec{r}_s} m_\phi(\vec{r}_s) \quad (5)$$

$$m_\phi(\vec{r}_s) \equiv \delta(\hat{x}_\phi \cdot \vec{r}_s) u(\hat{z}_\phi \cdot \vec{r}_s) \quad (6)$$

where \hat{x}_ϕ and \hat{z}_ϕ are the unit vectors in a rotated coordinate system (see figure 1(b)), $u(z)$ is the unit step (Heaviside) function, and $\delta(x)$ is a one-dimensional Dirac delta function. This form is useful as it simplifies analytical interpretation and is suggestive of numerical optimizations. Analytically, $m_\phi(\vec{r}_s)$ can be written in terms of the $\phi = 0$ case via an introduction of a rotation matrix:

$$m_\phi(\vec{r}_s) = m_0(R_\phi \vec{r}_s) \quad (7)$$

$$m_0(\vec{r}_s) = u(z) \delta(x) \quad (8)$$

$$R_\phi = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}. \quad (9)$$

This manipulation demonstrates that in accordance with the straight ray model, the one-way time of flight is given by the convolution of the slowness perturbation with a rotated, laterally infinitesimal unit step function, and provides a convenient method of generating the convolutional kernel for arbitrary angle ϕ . Thus the total lag map estimated by cross-correlation can be modeled as a convolution of the slowness perturbation and an impulse response $h(\vec{r}; \alpha, \beta)$:

$$\Delta\tau(\vec{r}_s; \alpha, \beta) = \sigma(\vec{r}_s) *_{\vec{r}_s} h(\vec{r}_s; \alpha, \beta) \quad (10)$$

$$h(\vec{r}_s; \alpha, \beta) = [m_0(R_\alpha \vec{r}_s) + m_0(R_{-\alpha} \vec{r}_s)] - [m_0(R_\beta \vec{r}_s) + m_0(R_{-\beta} \vec{r}_s)]. \quad (11)$$

A discretization of this impulse response is rendered in figure 2. Note the impulse response is composed of four lines of differing polarity. When this impulse response is convolved with the slowness, characteristic *phase shadow* signal features manifest in the lag map data below any sound speed inclusions. The polarity of these phase shadow features is dictated by the choice of reference signal in the cross-correlation processing.

2.2. Fourier space coverage

Because the forward operator which maps slowness to lag maps is a convolution, its Fourier space (Walker and Trahey 1998) coverage can be easily computed. As noted above, the convolutional kernel $m_\phi(\vec{r})$ can be written as a rotation of the unrotated kernel $m_0(\vec{r})$. Making use of the rotation property of the Fourier transform (Blahut 2004, Gonzalez and Woods 2008)

$$S(\vec{f}) = \mathcal{F}\{s(\vec{r})\} \implies \mathcal{F}\{s(R_\phi \vec{r})\} = S(R_\phi \vec{f}) \quad (12)$$

and noting that

$$M_0(\vec{f}) = \mathcal{F}\{u(z) \delta(x)\} = \frac{1}{2} \left(\frac{1}{j\pi f_z} + \delta(f_z) \right) \quad (13)$$

we have

$$M_\phi(\vec{f}) = \frac{1}{2} \left(\frac{1}{j\pi \hat{z}_\phi \cdot \vec{f}} + \delta(\hat{z}_\phi \cdot \vec{f}) \right). \quad (14)$$

It is also interesting to relate this approach to traditional transmission time of flight tomography (Greenleaf *et al* 1975, Glover and Sharp 1977). In the transmission tomography scenario, detectors are typically oriented opposite from the transmitters and the time of flight estimates to each detector are recorded. However, because

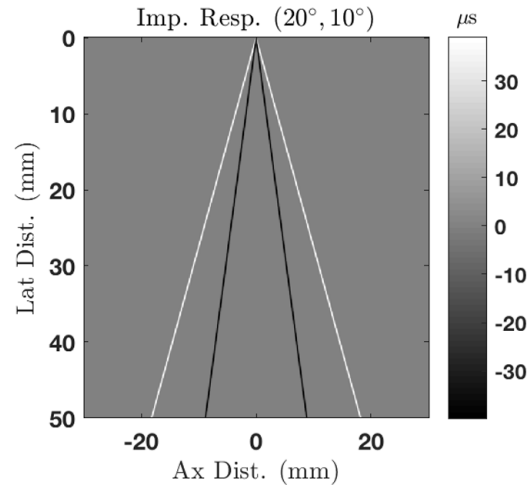


Figure 2. The impulse response of the straight ray model evaluated for the angles 20° and 10° . The model is composed of four terms which correspond to positive and negative lines in the data. Note that these lines are dependent on both the transmit and receive steering angles of both of the post beamformed frames, which are assumed to be symmetrically coupled (i.e. $\phi_{Rx} = -\phi_{Tx}$). The polarity of each path is determined by convention used in cross-correlation.

the data can only be collected at the receiver locations, less information is recorded than in the pulse-echo case and the form of the integral relation reduced down to a traditional Radon transform. Because the correlation-based approach allows for data to be estimated throughout the sample, additional depth-dependent information can be inferred. This results in a generalization of the Fourier projection-slice theorem, which typically only accounts for the delta function term. The resulting $1/j\pi\hat{\mathbf{z}}_\phi \cdot \vec{\mathbf{f}}$ decay is strictly the result of the depth dependent data encoded in the lag maps.

Generalizing the result in equation (14) to the four term result in equation (11), we have

$$\Delta\tau(\vec{\mathbf{f}}; \alpha, \beta) = \sigma(\vec{\mathbf{f}})H(\vec{\mathbf{f}}; \alpha, \beta) \quad (15)$$

$$H(\vec{\mathbf{f}}; \alpha, \beta) = \frac{1}{2} \left[\frac{1}{j\pi\hat{\mathbf{z}}_\alpha \cdot \vec{\mathbf{f}}} + \delta(\hat{\mathbf{z}}_\alpha \cdot \vec{\mathbf{f}}) + \frac{1}{j\pi\hat{\mathbf{z}}_{-\alpha} \cdot \vec{\mathbf{f}}} + \delta(\hat{\mathbf{z}}_{-\alpha} \cdot \vec{\mathbf{f}}) - \frac{1}{j\pi\hat{\mathbf{z}}_\beta \cdot \vec{\mathbf{f}}} - \delta(\hat{\mathbf{z}}_\beta \cdot \vec{\mathbf{f}}) - \frac{1}{j\pi\hat{\mathbf{z}}_{-\beta} \cdot \vec{\mathbf{f}}} - \delta(\hat{\mathbf{z}}_{-\beta} \cdot \vec{\mathbf{f}}) \right]. \quad (16)$$

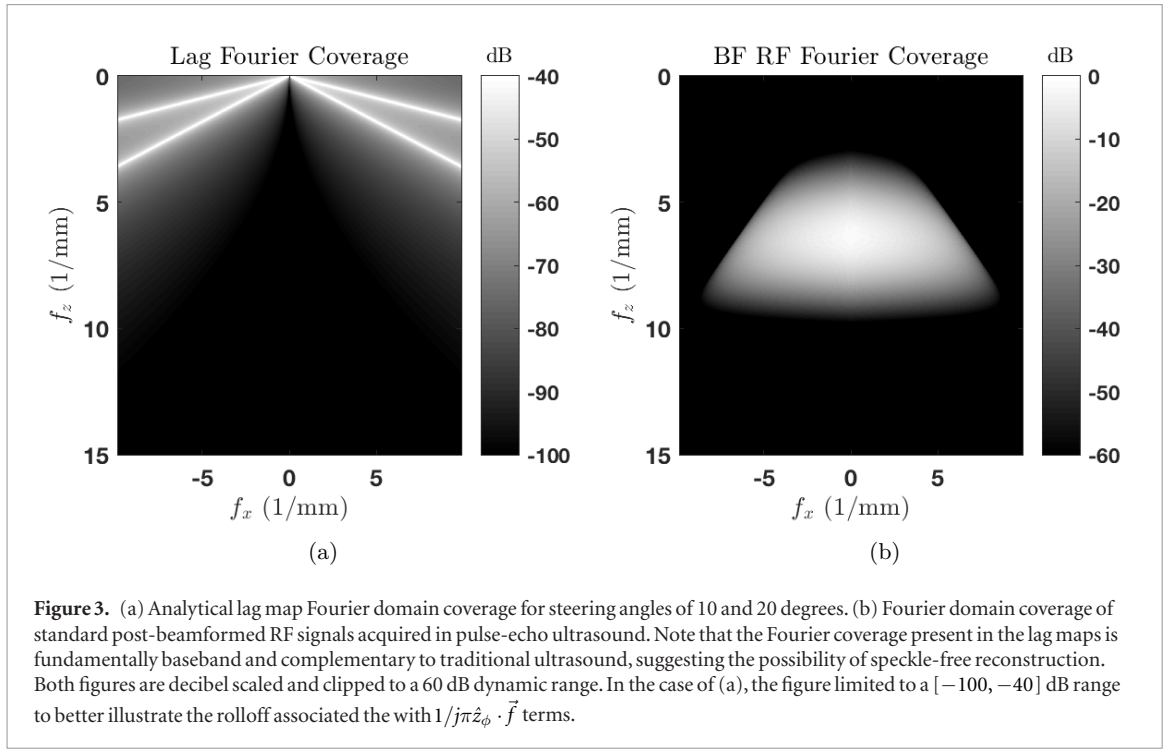
This spatial frequency coverage is rendered in figure 3(a). For comparative purposes, the spatial Fourier coverage for traditional pulse-echo beamformation (Walker and Trahey 1998) is rendered in figure 3(b) as given by

$$S(\vec{\mathbf{f}}, z) = \frac{\chi(f_t)}{f_z^2} E\left(\frac{cf_z}{2}\right) \left[A_T\left(\frac{-2z}{f_z}f_x\right) *_{f_x} A_R\left(\frac{-2z}{f_z}f_x\right) \right], \quad (17)$$

where $\chi(f_t)$ is the frequency dependent reflectivity, $E(f_t)$ is the temporal Fourier transform of the excitation signal¹, and $A_T(x)$ and $A_R(x)$ are the transmit and receive aperture functions respectively. Such a model can also be thought of as a proxy for a simple diffraction tomographic reconstruction of sound speed under a first Born scattering model (Norton and Linzer 1981, Simonetti and Huang 2008, Devaney 2012), because variations in sound speed manifest as reflectivity targets in pulse-echo mode. Of course, diffraction tomographic reconstructions of sound speed can be incorrectly interpreted, as density variations could produce identical signals, but the fact remains that information about sound speed exists at these higher frequency components.

To facilitate comparisons between the Fourier domain coverage of the lag maps and the traditional backscattered RF, the Fourier coverage of a 5 MHz, 128-element linear array (L9-4/38, Ultrasonix, Peabody MA) was computed using equation (17) with full apertures on transmit and receive and rendered in figure 3(b). Note that in the case of traditional post-beamformed ultrasound, the acquired data are bandpass in nature, which results in the characteristic speckle that dominates most B-mode ultrasound images. In contrast, the data acquired from the lag maps are fundamentally baseband in nature, suggesting the possibility of speckle-free reconstructions, albeit with lower spatial resolution.

¹ Convolved with the element electromechanical impulse response on both transmit and receive, which is assumed to be the same for each element.



Moreover, the angles of the phase lag lines can be manipulated by adjusting the steering angles, allowing for more diverse Fourier data to be acquired with multiple acquisitions. With multiple lag maps, the quantitative inverse problem can be viewed as inverting the data from a partially correlated *filter bank*, which allows for more data to be exploited than the single filter case. The details of such a multi-lag map reconstruction algorithm are explored in the following section.

2.3. Implications for reconstruction

The convolutional interpretation of the straight ray model yields a number of insights, which are particularly useful in development of a reconstruction algorithm. Because the forward operator that describes the lag maps is convolutional in nature, reconstruction is essentially a deconvolution problem. However, unlike many deconvolution problems studied in imaging, the convolutional kernel $h(\vec{r}; \alpha, \beta)$ has an infinite impulse response (IIR). This means that for any arbitrary $\sigma(\vec{r})$ of finite support, the lag map $\Delta\tau(\vec{r}; \alpha, \beta)$ spans all of \mathbb{R}^2 . Given that in practice the field of view is limited, the recorded data is in effect a *truncation* of the lag map, such that

$$\Delta\tau(\vec{r}; \alpha, \beta) = w(\vec{r}) \left[h(\vec{r}; \alpha, \beta) *_{\vec{r}} \sigma(\vec{r}) \right] \quad (18)$$

where $w(\vec{r})$ is a windowing function characterizing the support of the data. This windowing operation means that the forward operator is actually shift-variant, despite the convolutional form of the model.

For the purposes of this work, we will direct our attention to a Tikhonov-regularized least squares reconstruction (Tikhonov *et al* 1977), which solves the problem:

$$\vec{\sigma}_e = \underset{\vec{\sigma}}{\operatorname{argmin}} |\Delta\vec{\tau} - \mathbf{H}\vec{\sigma}|^2 + \gamma^2 |\vec{\sigma}|^2, \quad (19)$$

where $\vec{\sigma}$ and $\vec{\sigma}_e$ are the vectorized slowness decision variable and estimate respectively, $\Delta\vec{\tau}$ is the vectorized lag map data, \mathbf{H} is the forward model, and γ is the regularization hyperparameter used to compensate for operator ill-conditioning. In this context, the fact that the forward operator is a truncated convolution does lend itself to some useful consequences. In particular, it admits the possibility of matrix-free reconstruction algorithms, which are well known to have significant storage and computational benefits (Diamond and Boyd 2016). Because of the convolutional nature of the forward operator, the forward, adjoint, and normal operators can be replaced with equivalent matrix-free subroutines by exploiting algorithms such as FFT convolution (Oppenheim and Schaffer 1999). In conjunction with a conjugate gradient based inverse solver, such matrix-free subroutines can enable matrix-free reconstruction. To this end, we propose to use the biconjugate-gradient stabilized (BiCGStab) algorithm (van der Vorst 1992, Barrett *et al* 1995, Lin and Chew 1996), because it is unconditionally stable and is natively supported in MATLAB using matrix-free subroutines. In avoiding explicit matrix construction during the reconstruction process, we reduce the storage requirements for an $M \times N$ pixel reconstruction from $\mathcal{O}((MN)^2)$ values to $\mathcal{O}(MN)$ complex values, which is a better target for local reconstruction on modern clinical machines.

In order to reformulate this into a form more appropriate for BiCGStab, we note that the problem can be rewritten as $\vec{\sigma}_e = \mathbf{R}_H^{-1}(\gamma)\vec{\mathbf{d}}$, where

$$\mathbf{R}_H(\gamma) \equiv \mathbf{H}^* \mathbf{H} + \gamma^2 \mathbf{I} \quad (20)$$

$$\vec{\mathbf{d}} \equiv \mathbf{H}^* \Delta \vec{\tau}. \quad (21)$$

This is equivalent to solving $\mathbf{R}_H(\gamma)\vec{\sigma} = \vec{\mathbf{d}}$, and consequentially all that is needed is a subroutine to evaluate $\mathbf{R}_H(\gamma)\vec{\sigma}$, which will be run once per iteration of the BiCGStab algorithm.

For the case of multiple lag maps, we describe the total forward operator \mathbf{H} as a ‘vertical concatenation’ of individual operators \mathbf{H}_i as described in equation (7) of Jaeger *et al* (2015a):

$$\begin{bmatrix} \Delta \vec{\tau}_1 \\ \vdots \\ \Delta \vec{\tau}_N \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_N \end{bmatrix} \vec{\sigma} \triangleq \mathbf{H} \vec{\sigma}. \quad (22)$$

With analogy to equations (20) and (21), the BiCGStab algorithm may be implemented by deriving matrix-free routines for $\mathbf{R}_H(\gamma)$ and \mathbf{H}^* . Fortunately, this can be accomplished via a straightforward generalization of the single lag map case, because

$$\mathbf{H}^* \{\Delta \vec{\tau}\} = \sum_{i=1}^N \mathbf{H}_i^* \{\Delta \vec{\tau}_i\}, \quad (23)$$

$$\mathbf{R}_H(\gamma) \{\vec{\sigma}\} = \sum_{i=1}^N \mathbf{H}_i^* \{\mathbf{H}_i \{\vec{\sigma}\}\} + \gamma^2 \vec{\sigma}. \quad (24)$$

The matrix-free evaluation of $\gamma^2 \mathbf{I} \vec{\sigma}$ is straightforward, and a matrix-free subroutine for the regularized normal operator $\mathbf{R}_H(\gamma)$ can be written as:

Algorithm 1. Matrix-free evaluation of $\mathbf{R}_H(\gamma)\vec{\sigma}$.

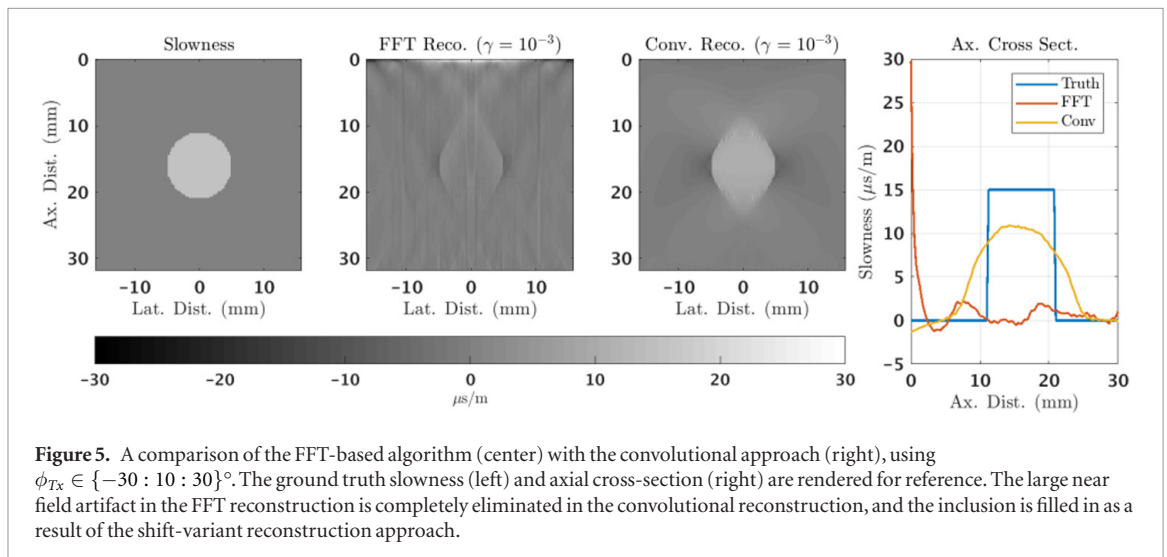
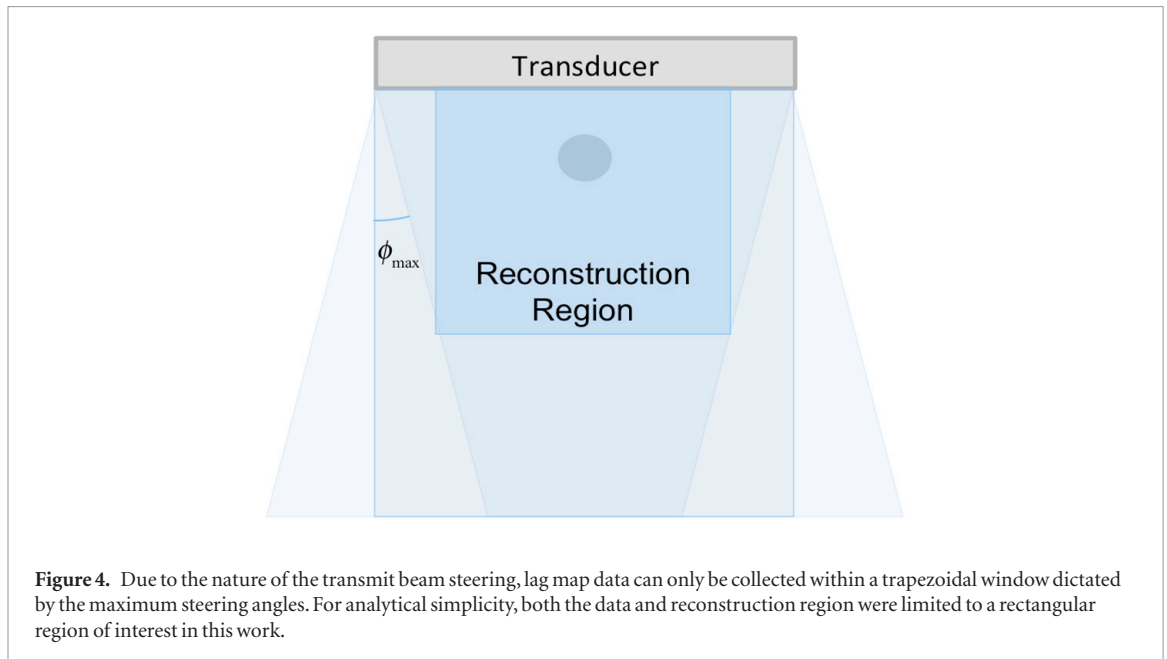
```

1: for all lag maps  $\Delta \vec{\tau}_i$  do
2:   Evaluate  $\Delta \vec{\tau}_i = \mathbf{H}_i \vec{\sigma}$  by FFT convolution.
3:   Window to data size.
4:   Evaluate  $\vec{\sigma}_i = \mathbf{H}_i^* \Delta \vec{\tau}_i$  by FFT convolution.
5:   Window to reconstruction region.
6: end for
7: Aggregate and add in the regularization term:  $\vec{\sigma}' = \left( \sum_i \vec{\sigma}_i \right) + \gamma^2 \vec{\sigma}$ .
```

Using these expressions, each iteration of BiCGStab will require $2N$ truncated convolutions, which is expensive given the fact that the impulse responses need to be at least the size of the reconstruction to minimize edge effects. There is therefore an incentive to reduce the total number of lag maps acquired while retaining as much quality as possible. Assuming the maximum steering angle is already specified due to field of view constraints (see figure 4), such a situation naturally lends itself to sparser acquisitions in terms of steering angles. However, sparsification of the angles may result in decorrelation of the post-beamformed data due to refractive errors, so the choice of an appropriate angular step remains a nuanced trade-off.

One could overlook the influence of steps 3 and 5 in the reconstruction, and directly invert in the Fourier domain under the presumption that it is approximately diagonal. Such an approach would only be optimal if the forward operator were a circular convolution, which is not true in practice. The distinctions between convolution and circular convolution are explicitly described in Diamond and Boyd (2016), as well as many textbooks. For this reason, any such reconstructions will be plagued by FFT wraparound artifacts associated with the circular convolution operation, whereas such effects will be avoided using our method. An illustration of this effect is shown in figure 5, where artifacts are present at the top of the reconstruction. To control for the effects of regularization in this example, the regularization parameter was held at a constant value of $\gamma = 10^{-3}$ for both reconstructions.

At this time, it is appropriate to contrast this method with the FFT-based approach of Jaeger and Frenz 2015, Jaeger *et al* 2015a, 2015b. As written, the algorithm describes a linear relationship between the DFT's of the lags and slowness which is implemented as a matrix multiplication. Based on the text of Jaeger *et al* (2015a), the main motivation for doing so is to exploit a Kronecker delta function in the operator associated with the lateral spatial frequencies, which was reported to strongly reduce the numerical complexity in terms of the inversion. Under



these conditions, the proposed operator would have a block diagonal structure in matrix form, and would require storage of $\mathcal{O}(M^2N)$ complex values per lag map, assuming sparse matrix storage. However, the particulars of this method were difficult to follow due to a number of errors, both explicit (dropping terms) and implicit (conflating the discrete Fourier transform with the continuous Fourier transform, and delta functions with sinc and Dirichlet functions), and as such the method does not appear to work as written. As such, we have failed to completely reproduce the results of that study.

Despite these difficulties, it would appear that Sanabria *et al* has been successful in implementing Jaeger's technique (Jaeger *et al* 2015a), and compared it to their spatial domain approach (Sanabria *et al* 2018). Our method would be most similar to the variant described in figure 4(b) in that paper, in which Tikhonov regularization was used with a spatial domain operator under straight ray modeling conditions. Under equivalent windowing and lateral antialiasing conditions, the methods would be identical in effect, though obviously differ in implementation details. However, comparisons with this variant were not performed on experimental or full wave simulated data, as they were more interested in studying their anisotropically weighted total variation method under those conditions. As such, one could not reasonably expect the reported quality of those reconstructions to generalize under realistic conditions. Furthermore, no mention of the convolutional relationship was made within that paper, suggesting that the authors used full-matrix methods, and were unaware of the theoretical and storage benefits of the convolutional formalism. Consequently, they would have used the full $\mathcal{O}((MN)^2)$ storage in implementation. Given the fact that such methods tend to be solved iteratively, it is likely that this method would benefit from the usage of matrix-free convolution algorithms, at least from a storage perspective.

In this paper, the Tikhonov-regularized spatial domain reconstruction will be studied in greater detail under full-wave conditions in both simulation and phantom experiments, as will be shown in section 3.

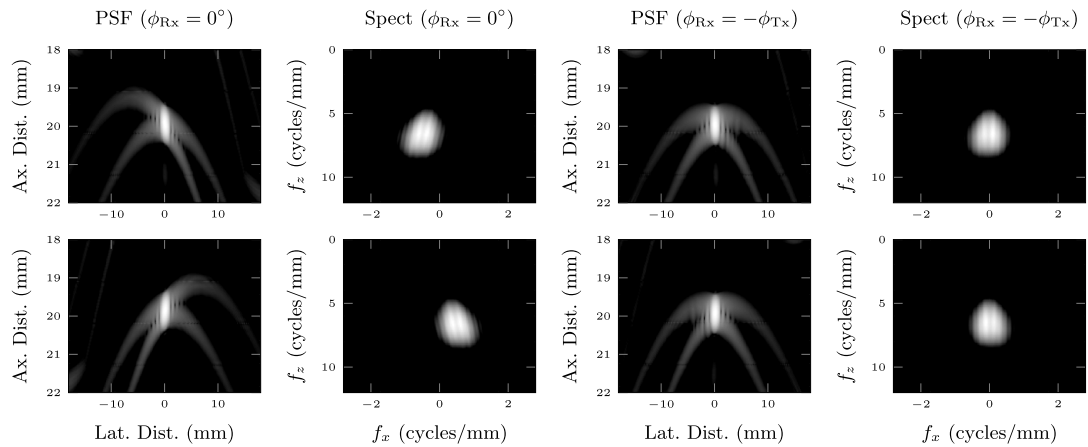


Figure 6. Point spread functions and 2D spectra of the post-beamformed RF under broadside receive apertures (left) and mirrored receive apertures (right). The top row corresponds to a transmit angle of -10° and the bottom row corresponds to a transmit angle of 10° . Observe that if we use a broadside aperture, the PSF and spectra rotate as a function of the transmit angle, whereas using a mirrored aperture keeps the spectra invariant.

2.4. Beamformation considerations

One particular difference from the prior published work in this domain is the fact that this approach takes into account the effect of the receive aperture in the beamformation process. Conventional wisdom in plane wave beamformation is to use a full receive aperture for the beamformation process, as this can achieve optimal spatial resolution after coherent compounding (Montaldo *et al* 2009). However, using a full receive aperture results in a point spread function which varies with position, which can be observed by rendering a B-mode image of an array of point targets. This shift variance makes it difficult to analyze the cross-correlation of the post-beamformed RF data spectrally, and consequentially difficult to analyze the effect of receive path on the lag map data.

As noted in Lu *et al* (1994), one way to achieve an approximately position-independent point spread function (PSF) is to use a dynamic receive focused aperture that preserves a constant F-number criterion, where the F-number is defined as the ratio of reconstruction depth to aperture width. Under this condition, the most natural implementation would be to use a receive aperture centered around the desired point of reconstruction, a method which we call broadside receive beamformation (i.e. $\phi_{Rx} = 0^\circ$). The PSF for this method is rendered on the left of figure 6. Observe that the PSF and the corresponding spectra rotate as a function of transmit steering angle. In the case of the coherent plane wave compounding scenario, this phenomenon is advantageous, as it allows for a higher resolution reconstruction due to the additional lateral frequency components acquired from the transmit steering process. However, for the purposes of extracting lag map data, this results in a decorrelation of the signals, as the spectral overlap is minimal. One should note that such error can also manifest with a full receive aperture as well, though the effect will depend on position. In the case of Jaeger *et al* (2015a), the transducer was also mechanically translated, which may have also exacerbated this effect.

In order to keep the spectra of the acquisitions approximately invariant to transmit angle, it is advantageous to use a receive aperture centered at an angle of $\phi_{Rx} = -\phi_{Tx}$, a condition we call a mirrored receive aperture. As shown on the right side of figure 6, using a mirrored receive aperture prevents the PSF and spectrum from rotating, allowing for an improved spectral overlap. This results in better correlation performance and consequently better lag map estimates, an important property which has been previously utilized in aberrating delay estimation (Rachlin 1990).

Since initial submission of this manuscript, the authors have become aware of an interesting preprint article from the Jaeger group which expands on this idea by utilizing a non-zero common mid-angle (Stähli *et al* 2019), which corresponds to the case when $\phi_{Tx} \neq -\phi_{Rx}$. This article supersedes many of the considerations made here, though still has yet to go through proper peer review. The generalization of this mirrored aperture approach to non-zero common mid-angles is particularly exciting due to a result that is predicted by the theory developed in section 2.2. With reference to equation (16) and figure 3, one can observe that the Fourier space coverage has a significant dip along the line $f_x = 0$. If the receive aperture is mirrored about a non-zero common mid-angle, this dip could be steered off the $f_x = 0$ line, thus giving access to more Fourier domain diversity than was available in the methods developed here, suggesting higher quality reconstructions. However, this claim regarding the Fourier domain coverage has yet to be studied in detail by our lab or reviewed by the research community at large, and at this point remains conjecture.

2.5. Simulation study

In order to assess the viability of the algorithm in a well-controlled environment, several simulations were conducted using the simulation package *k-wave*, with the relevant simulation parameters given in table 1. This package was chosen for its capability of modeling space-varying sound speed and density, and for its open source implementation. In each case, a 256-element 5 MHz linear array transducer was used to interrogate a medium with spatially varying sound speed and density distributions at a number of transmit steering angles. For each transmit event, the backscattered RF data were beamformed using a steered receive aperture diametrically opposed to the direction of propagation of the transmitted wave (i.e. $\phi_{Rx} = -\phi_{Tx}$) using spline interpolation and a reference sound speed value of $c_0 = 1540$ m/s. A temporal offset corresponding to the center of energy of the excitation was also used to keep the point spread function approximately invariant to transmit angle, as calculated by

$$\tau_c = \frac{\sum_{i=0}^{N_t-1} t_i e^2(t_i)}{\sum_{i=0}^{N_t-1} e^2(t_i)}, \quad (25)$$

where $e(t)$ is the excitation signal. The beamformation delay profiles used were the same as in Montaldo *et al* (2009), using spline interpolation to evaluate the signal at fractional sample lag values. The receive aperture grew dynamically to maintain a constant F-number criterion ($f/3$), in an attempt to preserve shift invariance of the point spread function as predicted by equation (17). This process is summarized explicitly below:

$$b(\vec{r}; \phi_{Tx}) = \sum_i p(t(\vec{r}, x_i; \phi_{Tx}) + \tau_c, x_i) A \left(\frac{f^\#}{z} (x_i - x - |\vec{r}| \sin \phi_{Rx}) \right) \quad (26)$$

where $p(t, x_i)$ is the backscattered RF signal at time instant t and element location x_i , $\vec{r} = [x; z]$ is the reconstruction location, $A(x_i)$ is the receive aperture, $f^\#$ is the f-number, and

$$t(\vec{r}, x_i; \phi_{Tx}) = \frac{1}{c_{bf}} \left(z \cos \phi_{Tx} + x \sin \phi_{Tx} + \sqrt{z^2 + (x - x_i)^2} \right). \quad (27)$$

The post-beamformed RF signals were converted into complex analytic signals $a(\vec{r}; \phi_{Tx})$ in the z direction using the FFT-based `hilbert()` function in MATLAB. Local time lag estimates of adjacent steering angle pairs were generated by the following formula

$$\Delta\tau(\vec{r}; \alpha, \beta) = \frac{1}{2\pi f_0} \text{atan}_2(\text{Im}\{R_a(\vec{r})\}, \text{Re}\{R_a(\vec{r})\}) \quad (28)$$

$$R_a(\vec{r}) = g(\vec{r}) *_{\vec{r}} [a(\vec{r}; \alpha) a^*(\vec{r}; \beta)] \quad (29)$$

where $g(\vec{r})$ is a tracking kernel. The authors used the outer product of squared Hamming windows for $g(\vec{r})$ based on empirical observations of efficacy, but such a choice should not be considered an optimal one. The resulting lag maps were then cropped down to a rectangular region within the field of view of the overlapping transmit beams to avoid artifacts within regions of poor SNR. Additionally, the data were downsampled using a spline downsampling filter in order to reduce both computational complexity and high frequency model-error noise, a choice which is justifiable given the low-pass nature of the forward model operator.

Note that because the analytic signal is evaluated in the z direction, the inherent meaning of $\Delta\tau(\vec{r}; \alpha, \beta)$ should not be interpreted as a true time-of-flight estimate, but rather an estimate of shift purely in the z -direction. For this reason, the model was modified to include a $\cos(\phi)$ scale factor for each term in the kernel, yielding the following form

$$h(\vec{r}; \alpha, \beta) = \cos(\alpha) [m_\alpha(\vec{r}) + m_{-\alpha}(\vec{r})] - \cos(\beta) [m_\beta(\vec{r}) + m_{-\beta}(\vec{r})]. \quad (30)$$

2.5.1. Effect of inclusion size

To analyze the effect of inclusion size on reconstruction, the radius of a circular inclusion was varied between 1.5, 3 and 4.5 mm respectively. The center of this inclusion was located at a depth of 10 mm centered in front of the transducer. This location was chosen because the method works better for near-field sound speed variations due to the greater availability of data, as predicted by the theory. The 1500 m/s inclusion was embedded in a 1540 m/s background medium, resulting in a corresponding slowness perturbation of $17.3 \mu\text{s m}^{-1}$. The corresponding density map $\rho(\vec{r})$ was generated by scaling the sound speed map by 1.5 and then adding 1% normally distributed random scattering $n(\vec{r}) \sim \mathcal{N}(0, 1)$ such that $\rho(\vec{r}) = c(\vec{r})/1.5 + (c_0/1.5)(0.01n(\vec{r}))$. The $n(\vec{r})$ term was incorporated to introduce weak diffuse scattering uniformly throughout the field of view for the purposes of tracking. The attenuation and nonlinearity features of the *k-wave* package were disabled for analytical simplicity.

No additive noise was added to the backscattered pressure data, in order to isolate the model-dependent effects from the statistical ones, as such statistical effects can be addressed by appropriate choice of regularization parameter.

2.5.2. Effect of inclusion contrast

To analyze the effects of inclusion contrast, the foreground sound speed was varied between 1500, 1540, and 1580 m/s for a 3 mm inclusion centered at a depth of 10 mm. Note that due to the nonlinear relationship between slowness and sound speed, the magnitude of the slowness perturbation ($\sigma \approx -16.4 \mu\text{s/m}$) for the 1580 m/s foreground ($\Delta c = 40 \text{ m/s}$) was lower than the 1500 m/s foreground sample ($\Delta c = -40 \text{ m/s}$, $\sigma \approx 17.3 \mu\text{s m}^{-1}$). The 1540 m/s (i.e. no contrast) simulation was included as a control for comparison purposes. All other settings were the same as in the previous simulation set.

2.6. Experimental validation

In order to validate this method empirically, a sound speed phantom was synthesized using an agar background and gelatin foreground material. The background material was synthesized by mixing 22.5 g noble agar (Sigma-Aldrich, St. Louis, MO) in 568 mL of degassed water, and adding 0.3 g of graphite powder (John Deere, Moline, IL) to induce scattering. The material was heated to approximately 80 °C in a microwave oven and allowed to cool to approximately 45 °C before pouring into a custom mold. The foreground material was similarly synthesized by combining 80 g gelatin (J.T. Baker, Phillipsburg, NJ) and 0.2 g graphite powder in 400 mL of degassed water, and heating to approximately 80 °C. The foreground material was poured into a 9.3 mm diameter cylindrical inclusion centered approximately 9.7 mm from the surface.

The sound speeds of the agar and gelatin materials were estimated to be 1491.2 m/s and 1542.0 m/s respectively via a pulse-echo insertion technique on homogeneous reference phantoms. To estimate these values, an unfocused 10 MHz, 0.375 inch diameter single element transducer (Valpey Fisher IP1001HR, Hopkinton, MA) was driven by a pulser-receiver (Panametrics 5800, Waltham, MA) and backscattered data were recorded by an A/D card. The phantom was suspended above a Plexiglas planar reflector in a water environment, and the reflection times associated with the front surface (t_f), back surface (t_b), and reflector (t_r) were estimated. After displacing the phantom out of the field of view, the reflection time to the planar reflector was estimated again (t_0). The sound speed in water c_w was derived from the temperature (Bilaniuk and Wong 1993), and the phantom sound speed c_p was calculated using the formula

$$c_p = c_w \left(\frac{t_0 - t_r}{t_b - t_f} + 1 \right). \quad (31)$$

The experimental parameters are given in table 2. The data were acquired using a 128-element array (Ultrasonix L9-4/38, Richmond BC, Canada) driven by a Verasonics Vantage 128, (Kirkland, WA). The prebeamformed RF were processed using the same beamformation algorithm used in section 2.5, using $c_0 = 1491.2 \text{ m/s}$ to compute the steering and beamformation delays. The receive f-number was increased to 5.715 (corresponding to 5° of angular coverage) to partially expand the field of view. The first 10 mm of lag map data was zeroed out due to reverberation clutter in the near field, and subsequently downsampled prior to reconstruction.

3. Results

3.1. Simulation study

Selected reconstructions for the inclusion size simulations are shown in figure 7(a). As predicted by the straight-ray model, the inclusion becomes increasingly apparent with increasing radius. In the low regularization regime, the inclusion is well-localized in the reconstruction. However, artifacts dominate the image at depths far away from the transducer surface. In the high regularization regime, the artifacts are suppressed, but the inclusion is blurred out and shifted upward, an effect which is even more apparent in figure 8. This phenomenon is expected due to the causal IIR nature of the forward model operator.

Laterally, the reconstruction is well resolved at all regularization levels as shown in figure 8, which is predicted from the Fourier description of the forward operator. Given the impulse response characterization of the forward operator, we do not expect to see any significant lateral degradation within the field of view, with the exception of boundary effects, which should begin at a distance of $|x| = z_{\max} \tan \phi_{\max} \approx 7 \text{ mm}$ away from the boundary according to the model. In principle, such boundary effects could be avoided by using a larger transducer, reducing the maximum angle or reducing the maximum depth. However, we chose to include them for visualization purposes.

In order to quantify performance, the root sum squared error (RSSE) of the slowness reconstructions are compiled in table 3. These values were computed using the formula

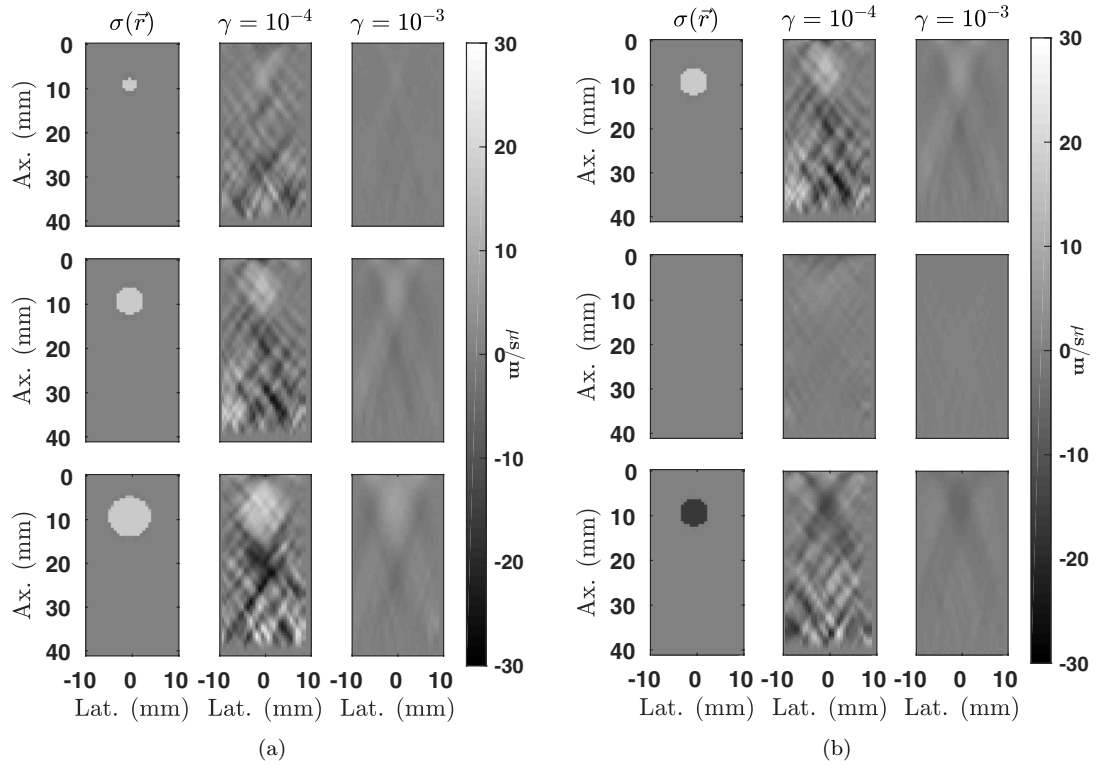


Figure 7. Slowness reconstructions for different size (a) and contrast (b) conditions at regularization levels of $\gamma \in \{10^{-4}, 10^{-3}\}$. In each case, the circular aberrating inclusion was located at a depth of 10 mm and the reconstruction region was limited laterally to ± 10 mm. In (a), the foreground slowness was set to a level of $17.32 \mu\text{s/m}$ ($\Delta c = -40 \text{ m/s}$, $c_0 = 1540 \text{ m/s}$), and the inclusion radii were 1.5, 3 and 4.5 mm respectively. In (b) the foreground slowness was given by $\sigma_{fg} \in \{17.3, 0, -16.4\} \mu\text{s/s}$ for each row respectively.

Table 1. k -wave simulation and reconstruction parameters.

Parameter	Value
k -wave simulation grid stepsize (Δz ; Δx ; Δt ; CFL)	(50 μm ; 76.2 μm ; 6.4935 ns; 0.2)
k -wave grid (N_z ; N_y ; N_t)	(824; 512; 9,889) samples
Perfectly matched layer (N_z ; N_x)	(74; 68) samples
Transducer pitch	152.4 μm
Excitation center frequency (f_0)	5 MHz
Excitation	10-cycle Gaussian tone burst
Steering angles (ϕ_{\min} ; $\Delta\phi$; ϕ_{\max})	(-10° ; 0.5° ; 10°)
Transmit apodization	Tukey cosine, 0.1 rolloff factor
Receive apodization	Kaiser, $\beta = 2$, $f/3$
Tracking kernel dimensions (L_z ; L_x)	(5.025; 2.5908) mm
Resampled phase lag grid step (Δz ; Δx)	(609.6; 609.6) μm
Phase lag map dimensions (N_z ; N_x)	(68; 30) samples
Stop criterion	10^{-10} relative residual or 3000 iterations

$$\text{RSSE} = |\vec{\sigma}(\gamma) - \vec{\sigma}_T|, \quad (32)$$

where $\vec{\sigma}(\gamma)$ is the vectorized slowness reconstruction at regularization parameter γ and $\vec{\sigma}_T$ is the ground truth slowness distribution. In general, these values increase with increasing radius, though as one can see in figure 7(a) the post-inclusion artifacts generally increase in amplitude as well. It should be noted that in choosing to evaluate the RSSE on slowness perturbation rather than sound speed, as has been done previously (Jaeger *et al* 2015a, Sanabria *et al* 2018), the effect of the nonlinear intensity transformation $c = 1/(\sigma + 1/c_{bf})$ is decoupled from the metric. Though this limits the interpretability of the metric, this choice makes the metric more sensitive to reconstruction errors, as $1/c_{bf} \approx 649 \mu\text{s m}^{-1}$ is more than an order of magnitude higher than any deviations in the rendered reconstructions and would consequently dominate the reported values of the RSSE. We have also chosen to use RSSE over the more popular root mean squared error here, as the root sum squared (RSS)

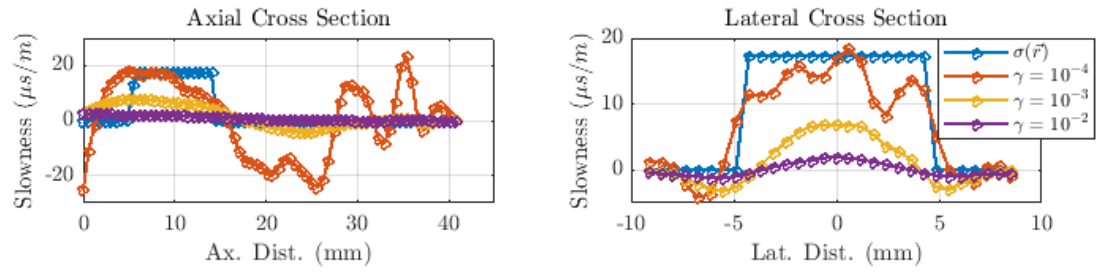


Figure 8. The axial (left) and lateral (right) cross-sections of the 4.5 mm radius reconstructions at various regularization parameters. The axial cross section was taken along the center line of the inclusion, the lateral cross section was taken at a depth of approximately 10 mm. As regularization increases, the reconstruction tends to smooth and damp at the reconstruction edges.

Table 2. Experimental acquisition and reconstruction parameters.

Parameter	Value
Transducer pitch	304.8 μm
Excitation center frequency (f_0)	4 MHz
Excitation	2-cycle tone burst
Transmit apodization	Tukey cosine, 0.1 rolloff factor
Receive apodization	Kaiser, $\beta = 2$, $f/5.715$
Correlation window dimensions (L_z ; L_x)	(3.97; 3.96) mm
Resampled phase lag grid step (Δz ; Δx)	(609.6; 609.6) μm
Phase lag map dimensions (N_z ; N_x)	(80; 44) samples
Stop criterion	10^{-10} relative residual or 10000 iterations

Table 3. Error performance metrics for the size sweep (top) and contrast sweep (bottom) experiments.

Radius (mm)	RSSE ($\mu\text{s m}^{-1}$)		Cosine Sim.	
	$\gamma = 10^{-4}$	$\gamma = 10^{-3}$	$\gamma = 10^{-4}$	$\gamma = 10^{-3}$
1.5	183.2	73.5	0.1415	0.2573
3	299.9	141.5	0.2205	0.3662
4.5	357.9	196.5	0.4125	0.4984

Δc (m/s)	RSSE ($\mu\text{s m}^{-1}$)		Cosine Sim.	
	$\gamma = 10^{-4}$	$\gamma = 10^{-3}$	$\gamma = 10^{-4}$	$\gamma = 10^{-3}$
-40	299.9	141.5	0.2823	0.5413
0	71.02	9.019	N/A	N/A
40	252.9	120.9	0.2205	0.3662

value of the ground truth would remain invariant under truncation and zero-padding transformations. The RSS value of the ground truth is an important quantity, as it represents the simple baseline of $\hat{\sigma} = 0$. Such a result can be interpreted as being equivalent to simply returning a sound speed value of c_{bf} after remapping. This baseline is important in the case of Tikhonov regularization, as it represents the high regularization limit of the reconstruction.

Because RSSE is not sensitive to changes in image shape, a cosine similarity metric was also evaluated, which is given by

$$\cos(\vec{\sigma}(\gamma), \vec{\sigma}_T) = \frac{\vec{\sigma}(\gamma)^* \vec{\sigma}_T}{|\vec{\sigma}(\gamma)| |\vec{\sigma}_T|}. \quad (33)$$

This metric can be thought of as the zero lag value of a normalized cross-correlation of the images, and thus is used as a metric because it evaluates the degree of registration of the image independent of its amplitude (Pinton *et al* 2006). Note that the cosine similarity is undefined for the zero contrast target, so those values were omitted in table 3.

To illustrate the intermediary steps in the process, examples from the 4.5 mm inclusion datasets are shown in figure 9. For analytical simplicity, we have restricted our analysis to the 5° and 5.5° transmit steering angles. By

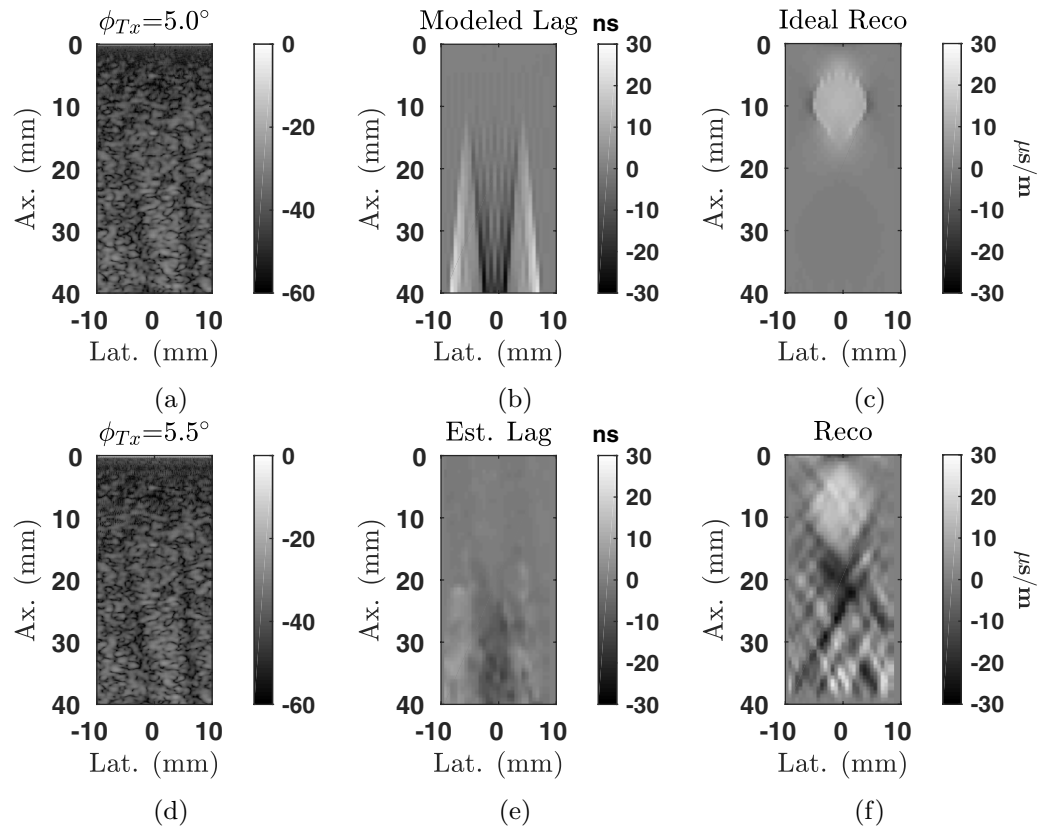


Figure 9. Sample images of the B-mode data (a) and (d) idealized and estimated lag maps (b), (e), and reconstructions ($\gamma = 10^{-4}$) (c) and (f) for the 4.5 mm inclusion dataset. Even though there is no discernible difference between B-mode frames for the 5° and 5.5° acquisitions, subtle phase shifts are detectable on the nanosecond scale. Despite an obvious model breakdown in the lag data, the inclusion is clearly localized in the reconstruction. Note that the artifacts under the inclusion are not present in the idealized reconstruction, suggesting that they are the result of higher order physics such as diffraction or refraction.

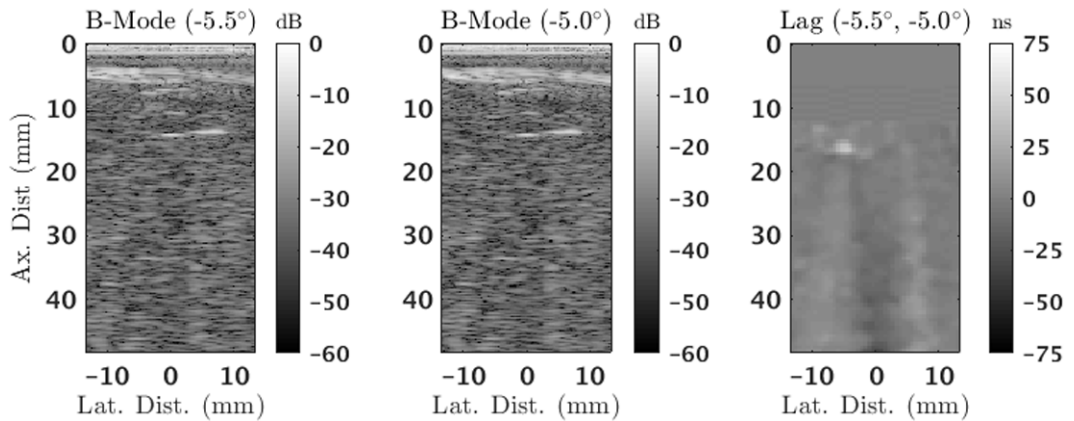
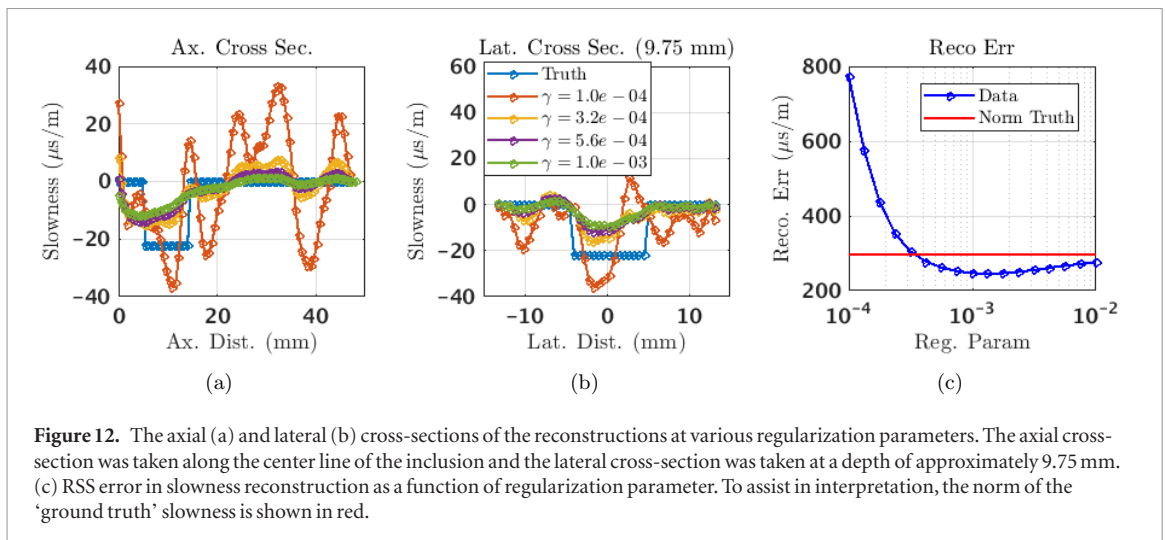
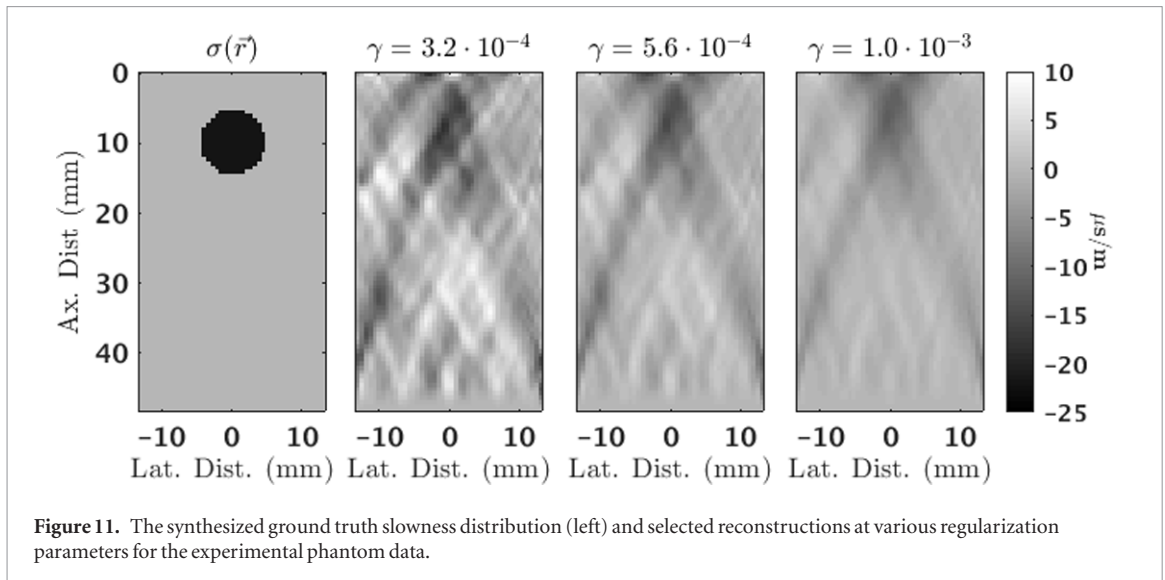


Figure 10. Example B-mode images (left, center) and associated lag map (right) from the phantom experiment. Note that the sound speed inclusion is not visible in the B-mode image, though reflections of the top and bottom of the inclusion can be identified. Due to the presence of reverberation error at a depth of approximately 5 mm, the lag maps are zeroed out above 10 mm.

comparing the B-mode frames, we observe no obvious differences in the post-beamformed datasets. However, subtle phase differences on the nanosecond scale can be observed in the lag map data, and the inclusion is clearly visible in the reconstruction. For comparative purposes, the reconstruction is compared to an idealized (inverse crime) reconstruction.

3.2. Experimental results

Selected reconstructions from the phantom experiment are rendered in figure 11, along with a synthesized ground truth image that was derived from the known *a priori* sound speed values. The inclusion location was determined



by the top reflection in the B-mode data. Example B-mode and lag map images are shown in figure 10. Much like the simulated reconstructions shown in section 3.1, the reconstructions are generally shifted slightly upward due to the depth-dependent damping effect of the regularization, which is more clearly observed in the cross-sections in figure 12. Generally, the magnitudes of the artifacts are somewhat larger in the experimental data than the simulated data when γ is held constant, and as a result the $\gamma = 10^{-4}$ reconstruction was omitted from figure 11.

In order to quantify the performance, the reconstruction error in slowness is plotted as a function of the regularization parameter in figure 12(c). For the purposes of this plot, the ground truth values $\vec{\sigma}_T$ were given by the synthesized distribution rendered in figure 11. As a reference point for quantitative comparison, the norm of $\vec{\sigma}_T$ is also plotted in red.

4. Discussion

By comparing the ideal reconstruction with the realistic reconstruction in figure 9, we may note two significant differences. First, the ideal reconstruction clearly has superior lateral resolution to the realistic reconstruction. It is likely that the idealized reconstruction is exploiting higher lateral frequency information that is not available in practice. This interpretation is justified by noting the reduction in spatial resolution between the estimated and modeled lag maps. It is possible some of these lateral frequency components can be recovered by incorporating beamwidth and tracking kernel size into the model. Unfortunately, because aberration causes the beam properties to degrade in practice, the utility of a beam correction will be limited in applicability, as the true beam will depend on the magnitude of the slowness perturbation.

Secondly, the artifacts in the realistic reconstruction are not present in the idealized reconstruction, which suggests that these artifacts are due to modeling error, most likely due to refraction and diffraction. Though k -wave can introduce numerical artifacts in the simulation process (i.e. pixelation in the 0–10 mm range of figure 9(d)) the continued presence of these artifacts in the experimental reconstructions in figure 11 suggests

that this is not the dominant source of error. Additionally, the reduction of the artifacts in the zero contrast reconstruction suggests that these artifacts are directly caused by the presence of the inclusion, which would be the case for refractive and diffractive errors. For the purposes of classifying these errors, we include misregistration of the positioning coordinates due to incorrect beamformation sound speed as refractive errors as well. Other potential sources of error include frequency dependence of scattering and noise in the phase lag estimates, but these effects would be present in the zero-contrast reconstructions as well, so either of these sources of error seems unlikely to explain these artifacts.

These artifacts are amplified by the fact that the deeper components of the reconstruction are more sensitive to perturbations than the upper components in the case of low regularization. This is due to the interplay between the truncation operation and causality constraint of the model forward operator. Specifically, because of the truncation of the lag maps to the data window, there are less data available which correspond to deeper targets than shallower targets. Consequentially, low regularization reconstruction operations tend to amplify these deeper targets, whereas the reconstruction can aggregate across a larger region of the lag map for shallower targets. This sensitivity explains why the $\gamma = 10^{-4}$ reconstructions have better cosine similarity values and subjective performance for the larger inclusions in figure 7(a) despite having worse quantitative error performance in table 3. It is likely that reducing the size of the reconstruction would result in better quantitative error performance by enforcing more data redundancy, though that is beyond the scope of this work.

Generally the artifacts in the real data reconstructions are slightly worse than in the simulations, to the point where the $\gamma = 10^{-4}$ reconstruction was incomprehensible. While there were several differences between the experimental and simulated data, perhaps the most significant one is the increased pitch of the array used in experiments. This is significant because the array used in experiments has the possibility of higher contribution from grating lobes. Though reducing excitation frequency to 4 MHz helps mitigate the grating lobe contribution, there could still be some partial grating lobe corruption in the RF above the center frequency. For example, the wavelength at 5 MHz is around $298.2 \mu\text{m}$, which is below the array pitch of $304.8 \mu\text{m}$, and thus even at broadside we expect some grating lobe corruption, however minor. Given that the received RF had spectral energy content at 5 MHz of around 6 dB below maximum, it is not out of the realm of possibility that high frequency grating lobes play a role in the lag map estimation process.

Another notable difference between the simulation and experimental data was the uniformity of the speckle. In the case of the simulation, the speckle was generated by an additive white Gaussian term to the density, whereas the experiment had no control for such things. A more sophisticated speckle tracking algorithm could possibly compensate for local variations in fast-time frequency (Loupas *et al* 1995), but that is beyond the scope of this paper.

As mentioned in section 2.3, the model is a truncated convolution, not a simple convolution. The introduction of the shift-variant truncation operation makes the reconstruction operator shift-variant as well. Because of this shift-variance, it is important to realize the limitations of a Fourier domain analysis. For example, the Fourier domain interpretation suggests that while the low frequency components are preserved in the lag maps, the DC bin² has low signal strength, due to the deconstructive interference of the individual $m_\phi(\vec{r})$ terms in the impulse response. Yet, the idealized reconstruction in figure 9(e) has a nonzero DC value.

This seemingly paradoxical result is explained by the shift-variance of the reconstruction operator. Much like how multiplying by a sinusoid in the time domain results in a frequency domain shifting (such as in some implementations of envelope detection), the introduction of a shift-variant reconstruction operator allows for reconstruction of out of band frequency components. Whether or not such components should be trusted is a matter of interpretation. A conservative interpretation of this phenomenon would reject the notion of DC slowness reconstruction based on the Fourier coverage alone. Here, the authors favor this interpretation, because the DC value of each individual lag map varied noticeably, indicating high noise in these estimates. In conjunction with the low DC gain of the individual operators, such an interpretation is justified. However, regardless of whether or not DC reconstruction can be trusted in principle, the low signal in the DC bin can explain the observed variability in layered phantom reconstructions such as those presented in Jakovljevic *et al* (2018).

More pragmatically, one must ultimately consider the convolutional model as an idealization. The data presented here could only be acquired in the presence of uniformly distributed scatterers. In practice, one would expect variations in echogenicity to be present in real world datasets, which limits the applicability of such an approach. This limitation is also shared in elastographic and Doppler imaging, though the inherent locality of the latter allows for some estimation robustness in the form of the power Doppler signal. Recently, this limitation was addressed through the introduction of space-varying regularization (Sanabria *et al* 2018), which may be ultimately required for any practical implementation. However, more research is required into alternative methods for dealing with this limitation.

² Direct current, from the electrical engineering nomenclature, referring to the global component of the data.

Furthermore, the overall linearity of the process should ultimately be called into question. Phase estimation via correlation methods is a nonlinear process, as evidenced by the arctangent in equation (28). In an attempt to make these estimates quasi-linear, the slowness was deliberately designed to produce small lags relative to $1/f_0$. In some situations, one would not have this luxury, and phase wrapping and other distortions could produce significant errors. Phase estimation methods that are robust to phase wrapping could be employed, but analytical interpretation of these methods is much more involved than what is presented here.

Finally, the modeling used in this work neglected the effects of frequency-dependent dispersion and nonlinearities. Though the effects of nonlinearities can be controlled in practice by keeping the incident pressure low, dispersion could result in both downshifting of the center frequency as well as introduce frequency dependence of sound speed, which was assumed to be invariant in this work. These effects should be considered when trying to reproduce these results outside of a numerical environment.

5. Conclusion

In this work, the model for registration-based local sound speed reconstruction was reformulated as a truncated convolutional operator, and some related theoretical consequences were explored. In particular, the model's Fourier space coverage was demonstrated to be largely orthogonal to the Fourier coverage of post-beamformed RF data utilized in traditional B-mode imaging. This is a significant finding because it suggests that the low spatial frequency components could be reconstructed with pulse-echo data, which is not recognized by pulse-echo reconstruction methods such as diffraction tomography. Furthermore, an iterative reconstruction algorithm was proposed based on this reformulation that allows for a dramatic reduction in memory storage requirements without introducing Fourier wraparound artifacts. More importantly, the imaging problem was directly related to the deconvolution problem, opening up the space of reconstruction algorithms to a large body of well-studied algorithmic approaches that can leverage the computational, storage, and theoretical properties of the convolution operation. The relationship to the Fourier projection-slice theorem of computed tomography was also explored. The performance of the reconstruction approach was tested in the open source simulation package *k-wave* and was characterized, and refractive and diffractive interpretations of the emerging artifacts were proposed. Additionally, a corresponding phantom experiment was performed and similar qualitative reconstruction performance was achieved.

Acknowledgments

The authors would like to acknowledge the contributions of Dr Michael Jaeger in the form of productive discussions of the practical considerations of his method, particularly the importance of receive beamformation steering angle and spatial domain reconstruction.

ORCID iDs

Anthony S Podkova  <https://orcid.org/0000-0001-6944-9066>

References

- Bamber J C and Hill C R 1981 Acoustic properties of normal and cancerous human liver-I. Dependence on pathological condition *Ultrasound Med. Biol.* **7** 121–33
- Bamber J C, Hill C R and King J A 1981 Acoustic properties of normal and cancerous human liver-II dependence on tissue structure *Ultrasound Med. Biol.* **7** 135–44
- Barrett R *et al* 1995 *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* vol 64 (Philadelphia, PA: SIAM)
- Bilaniuk N and Wong G S 1993 Speed of sound in pure water as a function of temperature *J. Acoust. Soc. Am.* **93** 1609–12
- Blahut R E 2004 *Theory of Remote Image Formation* 1st edn (Cambridge: Cambridge University Press) (<https://doi.org/10.1017/CBO9780511543418>)
- Byram B C, Trahey G E and Jensen J A 2012 A method for direct localized sound speed estimates using registered virtual detectors *Ultrason. Imaging* **34** 159–80
- Carson P L, Meyer C R, Scherzinger A L and Oughton T V 1981 Breast imaging in coronal planes with simultaneous pulse echo and transmission ultrasound *Science* **214** 1141–3
- Céspedes I, Ophir J and Huang Y 1992 On the feasibility of pulse-echo speed of sound estimation in small regions: simulation studies *Ultrasound Med. Biol.* **18** 283–91
- Chew W C 1995 *Waves and Fields in Inhomogeneous Media* 1st edn (New York, NY: IEEE)
- Devaney A J 2012 *Mathematical Foundations of Imaging, Tomography, and Wavefield Inversion* 1st edn (Cambridge: Cambridge University Press) (<https://doi.org/10.1017/CBO9781139047838>)
- Diamond S and Boyd S 2016 *Matrix-Free Convex Optimization Modeling* (Cham: Springer) pp 221–64
- Duric N, Littrup P, Poulou L, Babkin A, Pevzner R, Holsapple E, Rama O and Glide C 2007 Detection of breast cancer with ultrasound tomography: first results with the computed ultrasound risk evaluation (CURE) prototype *Med. Phys.* **34** 773–85

- Glover G H and Sharp J C 1977 Reconstruction of ultrasound propagation speed distributions in soft tissue: time-of-flight tomography *IEEE Trans. Son. Ultrason.* **24** 229–34
- Gonzalez R C and Woods R E 2008 *Digital Image Processing* 3rd edn (Upper Saddle River, NJ: Pearson Education, Inc.)
- Greenleaf J F, Johnson S A, Samayoa W F and Duck F A 1975 Algebraic reconstruction of spatial distributions of acoustic velocities in tissue from their time-of-flight profiles *Acoust. Hologr.* **6** 71–90
- Hormati A, Jovanović I, Roy O and Vetterli M 2010 Robust ultrasound travel-time tomography using the bent ray model *Proc. SPIE* **7629** 762901
- Huthwaite P and Simonetti F 2011 High-resolution imaging without iteration: a fast and robust method for breast ultrasound tomography *J. Acoust. Soc. Am.* **130** 1721–34
- Jaeger M and Frenz M 2015 Towards clinical computed ultrasound tomography in echo-mode: dynamic range artefact reduction *Ultrasonics* **62** 299–304
- Jaeger M, Held G, Peeters S, Preisser S, Grünig M and Frenz M 2015a Computed ultrasound tomography in echo mode for imaging speed of sound using pulse-echo sonography: proof of principle *Ultrasound Med. Biol.* **41** 235–50
- Jaeger M, Robinson E, Akarçay H G and Frenz M 2015b Full correction for spatially distributed speed-of-sound in echo ultrasound based on measuring aberration delays via transmit beam steering *Phys. Med. Biol.* **60** 4497–515
- Jakovljević M *et al* 2018 Local speed of sound estimation in tissue using pulse-echo ultrasound: model-based approach *J. Acoust. Soc. Am.* **144** 254–66
- Johnson S, Abbott T, Bell R, Berggren M, Borup D, Robinson D, Wiskin J, Olsen S and Hanover B 2007 Non-invasive breast tissue characterization using ultrasound speed and attenuation *Acoustic Imaging* vol 28 (Berlin: Springer) pp 147–54
- Kondo M, Takamizawa K, Hiram M, Okazaki K, Iinuma K and Takehara Y 1990 An evaluation of an *in vivo* local sound speed estimation technique by the crossed beam method *Ultrasound Med. Biol.* **16** 65–72
- Lavarello R and Oelze M 2008 A study on the reconstruction of moderate contrast targets using the distorted born iterative method *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **55** 112–24
- Lavarello R J and Oelze M L 2009 Tomographic reconstruction of three-dimensional volumes using the distorted born iterative method *IEEE Trans. Med. Eng.* **28** 1643–53
- Li S, Jackowski M, Dione D P, Varslot T, Staib L H and Mueller K 2010 Refraction corrected transmission ultrasound computed tomography for application in breast imaging *Med. Phys.* **37** 2233–46
- Lin J H and Chew W C 1996 BiCG-FFT T-matrix method for solving for the scattering solution from inhomogeneous bodies *IEEE Trans. Microw. Theory Tech.* **44** 1150–5
- Loupas T, Powers J and Gill R 1995 An axial velocity estimator for ultrasound blood flow imaging, based on a full evaluation of the Doppler equation by means of a two-dimensional autocorrelation approach *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **42** 672–88
- Lu J Y 1997 2D and 3D high frame rate imaging with limited diffraction beams *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **44** 839–56
- Lu J Y, Zou H and Greenleaf J F 1994 Biomedical ultrasound beam forming *Ultrasound Med. Biol.* **20** 403–28
- Mamou J and Oelze M L (ed) 2013 *Quantitative Ultrasound in Soft Tissues* (Dordrecht: Springer) (<https://doi.org/10.1007/978-94-007-6952-6>)
- Montaldo G, Tanter M, Bercoff J, Benech N and Fink M 2009 Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56** 489–506
- Norton S J and Linzer M 1981 Ultrasonic reflectivity imaging in three dimensions: exact inverse scattering solutions for plane, cylindrical, and spherical apertures *IEEE Trans. Biomed. Eng.* **28** 202–20
- Oelze M L and Mamou J 2016 Review of quantitative ultrasound: envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **63** 336–51
- Ophir J 1986 Estimation of the speed of ultrasound propagation in biological tissues: a beam-tracking method *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **33** 359–68
- Oppenheim A V and Schaffer R W 1999 *Discrete-Time Signal Processing* 2nd edn (Upper Saddle River, NJ: Prentice-Hall)
- Pinton G F, Dahl J J and Trahey G E 2006 Rapid tracking of small displacements with ultrasound *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53** 1103–16
- Rachlin D 1990 Direct estimation of aberrating delays in pulse-echo imaging systems *J. Acoust. Soc. Am.* **88** 191–8
- Sanabria S J, Ozkan E, Rominger M and Goksel O 2018 Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: simulation and *in vivo* study *Phys. Med. Biol.* **63** 215015
- Sehgal C M, Brown G M, Bahn R C and Greenleaf J F 1986 Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers *Ultrasound Med. Biol.* **12** 865–74
- Simonetti F and Huang L 2008 From beamforming to diffraction tomography *J. Appl. Phys.* **103** 103110
- Stähli P, Kuriakose M, Frenz M and Jaeger M 2019 Forward model for quantitative pulse-echo speed-of-sound imaging (arXiv:1902.10639)
- Tikhonov A N, Arsenin V I and John F 1977 *Solutions of Ill-Posed Problems* vol 14 (Washington, DC: Winston & Sons)
- Treeby B E, Jaros J, Rendell A P and Cox B 2012 Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a *k*-space pseudospectral method *J. Acoust. Soc. Am.* **131** 4324–36
- van der Vorst H 1992 Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems *SIAM J. Sci. Stat. Comput.* **13** 631–44
- Walker W F and Trahey G E 1998 The application of *k*-space in pulse echo ultrasound *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45** 541–58
- Wiskin J, Borup D, Johnson S, Berggren M, Abbott T and Hanover R 2007 Full-wave, non-linear, inverse scattering *Acoustical Imaging* (Berlin: Springer) pp 183–93