View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CiteSeerX

Geophys. J. Int. (2010) **180**, 1289–1302

Imaging From Sparse Measurements

Y. Fang, M. Cheney and S. Roecker
Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Accepted 2009 December 11; Received in original form 2009 April 7; doi: 10.1111/j.1365-246X.2009.04483.x

SUMMARY
We consider the problem of using scattered waves to recover an image of the medium in which the waves propagate. We address the case of scalar waves when the sources and receivers are sparse and irregularly spaced. Our approach is based on the single-scattering (Born) approximation and the generalized Radon transform. The key to handling sparse sources and receivers is the development of a data-weighting scheme that compensates for nonuniform sampling. To determine the appropriate weights, we formulate a criterion for measuring the optimality of the point-spread function, and solve the resulting optimization problem using regularized least-squares. Once the weights are determined, they can be used to compute the point-spread function and thus determine resolution, and they can also be applied to the measured data to form an image. Tests of our minimization scheme with different regularization parameters show that, with appropriate weighting, individual scatterers can be resolved at sub-wavelength scales even when data is noisy and the locations of both sources and receivers are uncertain. We show an example in which the source-receiver geometry and frequency bandwidths correspond to seismic imaging from multiple local earthquakes (passive seismic imaging). The example shows that the weights determined by our method improve the resolution relative to reconstructions with constant weights.

Key words: imaging – backprojection – sparse measurements.

1 INTRODUCTION

The analysis of records of scattered waves to recover images of the medium through which they propagate is a subject of great interest to many fields, including medical ultrasound imaging, radar imaging, and seismology. In general, the scattered wave inverse problem is nonlinear, but in many cases we can reasonably approximate the relationship between the medium and the scattered field by the *Born approximation* or *single-scattering* approximation. In this case the resulting linear operator is a *generalized Radon transform* (GRT) (e.g., Beylkin 1984; Beylkin 1985) which is a type of *Fourier integral operator* (FIO) (e.g., Duistermaat 1995; Grigis *et al.* 1994). An asymptotic inversion of the generalized Radon transform can be derived using FIO theory (e.g., Duistermaat 1995; Grigis *et al.* 1994). Just as for the classical *Radon transform* (e.g., Natterer *et al.* 2001), the inverse operator that maps scattered waves to the wave speed is a filtered adjoint operator of the forward operator. This method has the physical interpretation of *backprojection*.

The GRT approach reduces the inverse problem to one of inverting the Fourier-like transform from data generated and recorded at a certain set of points. When these points are uniformly sampled on a cartesian grid, the application of GRT is relatively straightforward. There are many instances, however, where either by design or physical necessity, the geometry is not so simple. For example, in radar imaging (e.g., Borden 1999), data is usually collected on a polar coordinate grid, and in magnetic resonance imaging (MRI) (e.g., Liang *et al.* 2000), the recorded data has a spiral sampling pattern. In the case of passive seismic source (e.g., earthquake) imaging with a regional array, both the source and receiver locations can be sparse and highly irregular.

To handle observations made on a non-cartesian grid, one can either interpolate to a cartesian grid, or attempt to handle the Fourier inversion directly. In the latter approach, ultimately a weight function needs to be chosen to compensate for the nonuniform sampling. In the literature of MRI, this weight function is referred to as a *sampling density compensation function*.

In the last few decades there has been abundant research related to the direct inversion of the Fourier transform on an irregular grid, much of it focused on the *Nonuniform Fast Fourier Transform* (NUFFT). The NUFFT arises not only in many inverse problems, but also in signal processing and image processing. Several studies (e.g., Greengard *et al.* 2004; Dutt *et al.* 1993; Nguyen *et al.* 1999; O'Sullivan 1985; Fessler *et al.* 2003) review methods using interpolation and gridding techniques that can achieve fast calculation of NUFFT with high accuracy, and other researchers (e.g., Jackson *et al.* 1991; Pipe *et al.* 1999; Malik *et al.* 2005) studied the sampling density weight function in the gridding method in MRI. Bydder *et al.* (2007) evaluated different density weight functions for the gridding method in MRI, and Samsonov *et al.* (2003) presented a method to calculate the sampling density weight function by optimizing the *point-spread function* based on the gridding method.

Interpolation, gridding, and related techniques (e.g., Wang 2003; Wang *et al.* 2009) are known to work well when the data is

densely sampled. However, in many seismic and radar imaging applications, sources and receivers are sparsely located, and in these cases, one can not apply NUFFT directly because the measured data is not simply a Fourier transform of the unknown medium. To convert the data to a Fourier transform, one must carry out additional steps of parameterizing the source-receiver geometry and changing the variables of integration.

We can forego these additional steps by avoiding interpolation and gridding altogether, and instead determining a weight function that not only compensates for the nonuniform sampling but also accounts for the Jacobian term from the change of variables. In this paper, we use this approach to address situations in which sources and receivers are sparsely located on an irregular grid. In particular, we derive a weight function that gives rise to an imaging system with the best point-spread function. To do this, we base our work on a mathematical model for the forward problem and then determine an inversion formula using the results from FIO theory. We then present an optimization method based on the notion of *generalized functions* (e.g., Friedlander *et al.* 1998), and discuss the intrinsic trade-off between resolution and stability. Finally, we show how these different weight functions may be applied in source-receiver geometries relevant to both active and passive source seismic imaging.

2 PROBLEM FORMULATION

The propagation of waves that emanate from an impulsive point source in an isotropic medium is governed, in the frequency domain, by the Helmholtz equation

$$\nabla^2_{\boldsymbol{x}} u(\boldsymbol{x},\boldsymbol{x}_s,\omega) + \frac{\omega^2}{v^2(\boldsymbol{x})} u(\boldsymbol{x},\boldsymbol{x}_s,\omega) = -\delta(\boldsymbol{x}-\boldsymbol{x}_s), \quad (1)$$

where $\boldsymbol{x} \in \mathbb{R}^3$ is the position in space, $\boldsymbol{x}_s \in \mathbb{R}^3$ is the source position, u is the displacement in frequency domain, ω is the angular frequency, $v(\boldsymbol{x})$ is the wave speed which depends on position, and $\delta(\boldsymbol{x})$ denotes the 3-dimensional Dirac delta function.

We consider the case in which the wave speed is the sum of two terms, a known smooth background speed $c(\boldsymbol{x})$ and a perturbation term that models scatterers or interfaces between different media. We write

$$\frac{1}{v^2(\boldsymbol{x})} = \frac{1}{c^2(\boldsymbol{x})}(1+\alpha(\boldsymbol{x})). \quad (2)$$

and we refer to $\alpha(\boldsymbol{x})$ as the perturbation.

The field $u(\boldsymbol{x},\boldsymbol{x}_s,\omega)$ from a point-like source can be written as the sum of an incident wave field $G_0(\boldsymbol{x},\boldsymbol{x}_s,\omega)$ and a scattered wave field $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$:

$$u(\boldsymbol{x},\boldsymbol{x}_s,\omega) = G_0(\boldsymbol{x},\boldsymbol{x}_s,\omega) + u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega), \quad (3)$$

where $G_0(\boldsymbol{x},\boldsymbol{x}_s,\omega)$ is the outgoing Green's function. The Green's function satisfies (1) with v replaced by the known background c.

We subtract the equation for G_0 from (1), substitute (2), and use Green's method (e.g., Jackson 1998) to solve for $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$. This results in the integral equation for the scattered field $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$:

$$u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega) = \omega^2 \int_{\mathbb{R}^3} \frac{\alpha(\boldsymbol{y})}{c^2(\boldsymbol{y})} G_0(\boldsymbol{x},\boldsymbol{y},\omega) \times \big(G_0(\boldsymbol{y},\boldsymbol{x}_s,\omega) + u^s(\boldsymbol{y},\boldsymbol{x}_s,\omega)\big) d\boldsymbol{y}. \quad (4)$$

Equation (4) is an integral equation that must be solved, rather than simply a formula for $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$. For the inverse problem,

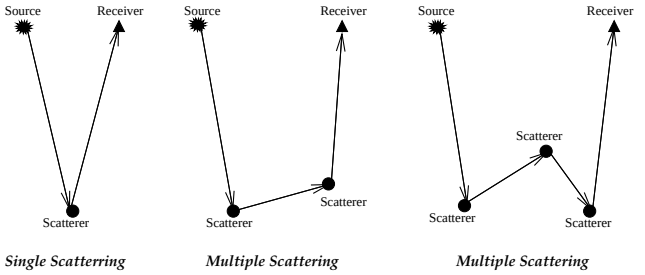


Figure 1. Examples of single and multiple scattering.

where we know $c(\boldsymbol{y})$, $G_0(\boldsymbol{x},\boldsymbol{y},\omega)$ and the values of $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$ at certain receivers, the right side of (4) contains a product of two unknown quantities: $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$ in the scattering region and $\alpha(\boldsymbol{x})$, which is the image of the medium we seek to determine. To linearize the problem, we use the Born approximation to neglect the $u^s(\boldsymbol{x},\boldsymbol{x}_s,\omega)$ term on the right side of (4). The Born approximation is valid when either the perturbation $\alpha(\boldsymbol{x})$ is small, or when most of the received energy is due to single scattering (Fig. 1). The Born approximation is thus

$$u^s_B(\boldsymbol{x},\boldsymbol{x}_s,\omega) = \omega^2 \int_{\mathbb{R}^3} \frac{\alpha(\boldsymbol{y})}{c^2(\boldsymbol{y})} G_0(\boldsymbol{x},\boldsymbol{y},\omega) G_0(\boldsymbol{y},\boldsymbol{x}_s,\omega) d\boldsymbol{y}. \quad (5)$$

Here the subscript 'B' stands for Born approximation.

In (5), we use the WKBJ approximation (e.g., Guillemin *et al.* 1977) for G_0:

$$G_0(\boldsymbol{x},\boldsymbol{y},\omega) = A(\boldsymbol{y},\boldsymbol{x}) e^{i\omega\tau(\boldsymbol{y},\boldsymbol{x})}, \quad (6)$$

where $\tau(\boldsymbol{y},\boldsymbol{x})$ is the travel time for the wave propagating from $\boldsymbol{y}$ to $\boldsymbol{x}$ and can be computed by solving the Eikonal equation (e.g., Bleistein *et al.* 2001). $A(\boldsymbol{y},\boldsymbol{x})$ is the amplitude that satisfies the first order transport equation (e.g., Bleistein *et al.* 2001).

In (5), we replace G_0 by (6) and replace $\boldsymbol{x}$ by $\boldsymbol{x}_r$ since data are collected only at locations of receivers ($\boldsymbol{x}_r$). We thus obtain an expression for the recorded data:

$$u^s_D(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega) = \omega^2 \int_{\mathbb{R}^3} \frac{\alpha(\boldsymbol{y})}{c^2(\boldsymbol{y})} a(\boldsymbol{x}_s,\boldsymbol{y},\boldsymbol{x}_r) e^{i\omega\phi(\boldsymbol{x}_s,\boldsymbol{y},\boldsymbol{x}_r)} d\boldsymbol{y}, \quad (7)$$

where the subscript 'D' stands for data. $a(\boldsymbol{x}_s,\boldsymbol{y},\boldsymbol{x})$ and $\phi(\boldsymbol{x}_s,\boldsymbol{y},\boldsymbol{x})$ are defined by

$$a(\boldsymbol{x}_s,\boldsymbol{y},\boldsymbol{x}_r) = A(\boldsymbol{x}_s,\boldsymbol{y}) A(\boldsymbol{y},\boldsymbol{x}_r), \quad (8)$$

$$\phi(\boldsymbol{x}_s,\boldsymbol{y},\boldsymbol{x}_r) = \tau(\boldsymbol{x}_s,\boldsymbol{y}) + \tau(\boldsymbol{y},\boldsymbol{x}_r). \quad (9)$$

2.1 Weighting Problem

The derivation of a inversion formula for the reconstruction problem in the ideal case where the data is known on a three dimensional set can be found in the appendix. In the case where sources $\boldsymbol{x}_s$ and receivers $\boldsymbol{x}_r$ are located on an irregular surface, we parametrize the source-receiver geometry $(\boldsymbol{x}_s,\boldsymbol{x}_r)$ by the two dimensional variable $\boldsymbol{\sigma} = (\sigma_1,\sigma_2)$, and use $\boldsymbol{\sigma}$ to represent $(\boldsymbol{x}_s,\boldsymbol{x}_r)$ in the remainder of the paper. To form an image, we use an approximation of the inversion formula (A.1)

$$I_W(\boldsymbol{x}) = \frac{1}{(2\pi\bar{c})^3} \sum_{\boldsymbol{\sigma}} W(\boldsymbol{x},\boldsymbol{\sigma}) \int \frac{c^2(\boldsymbol{x}) u^s_D(\boldsymbol{\sigma},\omega)}{a(\boldsymbol{x},\boldsymbol{\sigma})} e^{-i\omega\phi(\boldsymbol{x},\boldsymbol{\sigma})} d\omega \quad (10)$$

where $\boldsymbol{x}$ is an image point, ω is angular frequency, $I_W(\boldsymbol{x})$ is the image (subscript 'W' stands for weighted) of α, $W(\boldsymbol{x},\boldsymbol{\sigma})$ is the

unknown weight function, $\bar{c}$ is the average of the background speed, a is the amplitude (8), and ϕ is the total travel time (9). Here we scale the problem by $1/\bar{c}^3$ because each function ϕ in (A.13) is roughly of order $1/\bar{c}$ and we want to make the weight function a reasonable size.

We substitute the forward model for u_D^s (7) into (10) to obtain

$$I_W(\boldsymbol{x}) = \frac{1}{(2\pi\bar{c})^3}\sum_{\boldsymbol{\sigma}} W(\boldsymbol{x},\boldsymbol{\sigma})\int \frac{c^2(\boldsymbol{x})}{a(\boldsymbol{x},\boldsymbol{\sigma})}\omega^2 \times \int_{\mathbb{R}^3}\frac{\alpha(\boldsymbol{y})}{c^2(\boldsymbol{y})}a(\boldsymbol{y},\boldsymbol{\sigma})e^{-i\omega(\phi(\boldsymbol{x},\boldsymbol{\sigma})-\phi(\boldsymbol{y},\boldsymbol{\sigma}))}d\boldsymbol{y}d\omega. \quad (11)$$

In (11), we make the approximations [see also (A.8) and (A.9)]

$$\phi(\boldsymbol{x},\boldsymbol{\sigma})-\phi(\boldsymbol{y},\boldsymbol{\sigma}) \approx \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\cdot(\boldsymbol{x}-\boldsymbol{y}) \quad (12)$$

$$a(\boldsymbol{x},\boldsymbol{\sigma}) \approx a(\boldsymbol{y},\boldsymbol{\sigma}) \quad (13)$$

and the change of variables

$$(\omega,\boldsymbol{\sigma}) \mapsto \boldsymbol{k} = \omega\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma}), \quad (14)$$

to obtain

$$I_W(\boldsymbol{x}) \approx \int_{\mathbb{R}^3}\frac{1}{(2\pi\bar{c})^3}\sum_{\boldsymbol{\sigma}} W(\boldsymbol{x},\boldsymbol{\sigma})\int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\cdot(\boldsymbol{x}-\boldsymbol{y})}d\omega \times \alpha(\boldsymbol{y})d\boldsymbol{y} \quad (15)$$

$$= \int_{\mathbb{R}^3} K_W(\boldsymbol{x},\boldsymbol{y})\alpha(\boldsymbol{y})d\boldsymbol{y}, \quad (16)$$

where

$$K_W(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{(2\pi\bar{c})^3}\sum_{\boldsymbol{\sigma}} W(\boldsymbol{x},\boldsymbol{\sigma})\int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\cdot(\boldsymbol{x}-\boldsymbol{y})}d\omega \quad (17)$$

$$= \frac{1}{(2\pi\bar{c})^3}\sum_{\boldsymbol{x}_s,\boldsymbol{x}_r} W(\boldsymbol{x},\boldsymbol{x}_s,\boldsymbol{x}_r) \times \int \omega^2 e^{-i\omega(\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}_s,\boldsymbol{x})+\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x},\boldsymbol{x}_r))\cdot(\boldsymbol{x}-\boldsymbol{y})}d\omega. \quad (18)$$

Note that $W(\boldsymbol{x},\boldsymbol{\sigma}) = W(\boldsymbol{x},\boldsymbol{x}_s,\boldsymbol{x}_r)$. Both the travel times $\tau(\boldsymbol{x}_s,\boldsymbol{x})$ and $\tau(\boldsymbol{x},\boldsymbol{x}_r)$ satisfy the Eikonal equation. The vectors $\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}_s,\boldsymbol{x})$ and $\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x},\boldsymbol{x}_r)$ have the same length, namely $1/c(\boldsymbol{x})$, and are tangential to the corresponding ray paths at the point $\boldsymbol{x}$. The sum of these two vectors is a vector in the direction of the bisector of the angle between the vectors, and its length in general is between 0 and $2/c(\boldsymbol{x})$. Fig. 2 shows the bisector at one imaging point from one source-receiver pair in a homogeneous medium.

Given any imaging point $\boldsymbol{x}$, in order to form an image of α at $\boldsymbol{x}$, we would like to determine $W(\boldsymbol{x},\boldsymbol{\sigma})$ or $W(\boldsymbol{x},\boldsymbol{x}_s,\boldsymbol{x}_r)$ such that $K_W(\boldsymbol{x},\boldsymbol{y})$ is as close as possible to the Dirac delta function. The degree to which $K_W(\boldsymbol{x},\boldsymbol{y})$ approximates a delta function is determined by the source-receiver geometry and the frequency band of the data.

2.2 Weight Functions and Resolution

In the infinite band-width case, the set of bisectors at an imaging point characterizes the resolution that can be achieved at that point. In general, the directions in which most bisectors point will have the best resolution, while those with few bisectors have poor resolution. In the band-limited case (i.e., when ω is bounded), we define the *bisecting line segment*

$$\{\omega(\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}_s,\boldsymbol{x})+\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x},\boldsymbol{x}_r)) : \omega \text{ ranges over the measured angular frequencies}\} \quad (19)$$

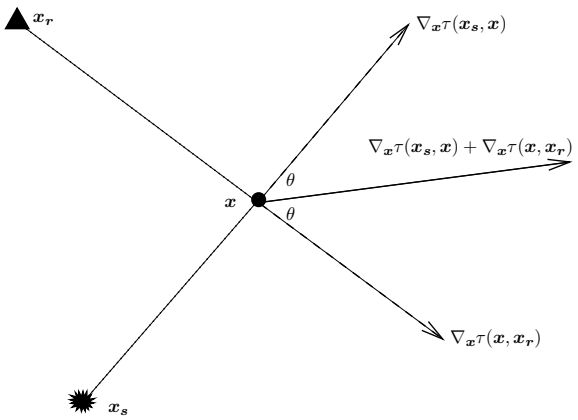


Figure 2. The bisector of one source-receiver pair in a homogeneous medium at imaging point $\boldsymbol{x}$.

as the bisector scaled by frequency. The set $\Xi_{\boldsymbol{x}}$ of points in the union of all the bisecting line segments determines the resolution of the backprojection. An example of a set $\Xi_{\boldsymbol{x}}$ at one imaging point is shown in Fig. 3.

A simple and intuitive choice for W is the constant weight function $W(\boldsymbol{x},\boldsymbol{x}_s,\boldsymbol{x}_r) = 1$. This is equivalent to calculating the average of the backprojected data from all the source-receiver pairs in reconstruction. A potential disadvantage of weighting each bisector equally is suggested by the simple example of $\Xi_{\boldsymbol{x}}$ in Fig. 3, in which all but one of the bisectors point in the horizontal direction. The point-spread function calculated with the constant weight function at that imaging point (Fig. 4(a)) consists of vertical ridges, which provide good resolution in the horizontal direction and almost no resolution in the vertical direction. In this case any information provided by the backprojected data from the source-receiver pair associated with the vertical bisector is overwhelmed by that from the horizontal bisectors. However, if we assign a larger weight value to the vertical bisector than to the horizontal one, the overall resolution is improved (Fig. 4(b)).

Another implication of this example is that constant weight functions can reduce the level of information when more observations are included in the inversion. Suppose initially we have only one source-receiver pair that gives a vertical bisector. This will provide some resolution in the vertical direction, but no resolution in the horizontal direction. If we add in source-receiver pairs with horizontal bisectors with a constant weight function, we gain resolution in the horizontal direction but lose resolution in the vertical direction. Because one would expect a reasonable weight function to allow additional data to improve an image reconstruction, it would appear that simple constant weights are not the best choice for irregular and sparse geometries. In the next section we discuss a variety of approaches to determining alternate weight functions.

3 DETERMINATION OF WEIGHT FUNCTIONS

3.1 Possible Approaches and Their Disadvantages

We investigated a number of approaches towards developing potential weight functions, and begin with a brief summary of those we found did not work well before extensively discussing one, the test-function approach, that did.

• *Backus-Gilbert method* (Backus *et al.* 1968): The Backus-Gilbert method is an inversion approach based on minimizing the

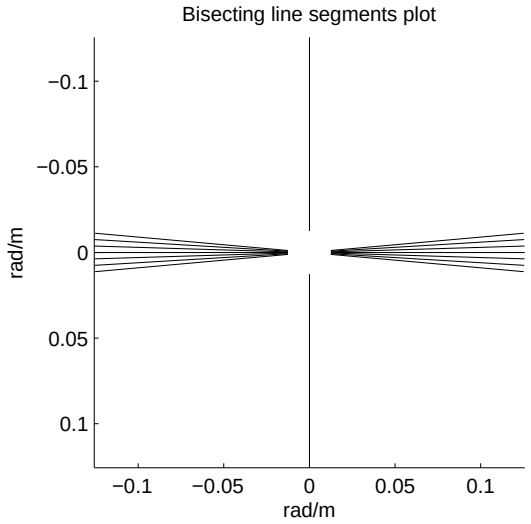
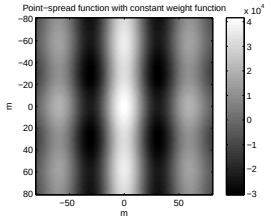
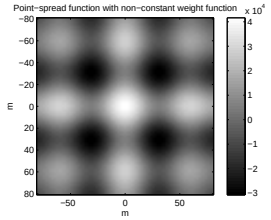


Figure 3. A 2-D example of $\Xi_{\boldsymbol{x}}$ with several near-horizontal bisectors and only one vertical bisector. The horizontal axis is k_1 and the vertical axis is k_2, where $\boldsymbol{k} = (k_1, k_2)$



(a) Point-spread function with constant weights. Horizontal axis is z_1 and vertical axis is z_2 where $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$ in equation (18). The units of point-spread function values are $1/m^2$.



(b) Point-spread function with non-constant weights. Axes are the same as in 4(a).

Figure 4.

variance of the point-spread function. This method works well in one dimensional problems but fails in higher dimensions due to the appearance of infinite variances in ridge-like terms that must be added together in constructing the point-spread function.

• *Interpolation and gridding-based methods*: Interpolation and gridding works well for applications such as MRI in which the sampling is nonuniform but dense, but are less robust when data sampling is sparse. Also, in order to use interpolation or gridding to evaluate integrals such as (A.16)), the first step is to parameterize the source-receiver geometry and compute (A.13), which is difficult in this case.

• The *Voronoi diagram* (e.g., Aurenhammer 1991): A Voronoi diagram, named after Georgy Voronoi, can be used to decompose a region into pieces based on distance to various points. For a problem in a bounded domain, a weight function can be derived from a Voronoi diagram. However, our problem involves an unbounded domain, and it is unclear what weights should be assigned to the Voronoi cells on the boundary of the computational domain. Moreover, using weights derived from a Voronoi diagram also requires first parameterizing the source-receiver geometry and computing (A.13).

• *Optimization using data as a constraint*: The data constraint approach expresses the desired perturbation α in terms of some basis, and minimizes the L^2 norm of α using the data as a constraint (e.g., Aster *et al.* 2005; Menke 1989). This method can be applied to the sparse measurement environment, but this is a different approach to the inverse problem. A disadvantage of this method lies in its dependence on the size of the data set, which could be very large. Moreover, any change in the data set requires that the entire minimization procedure be carried out again. Another disadvantage is that the size of the problem is determined by the number of voxels in the mesh of the computational domain, and hence it could be computationally expensive because a global solver is usually needed to compute the forward problem. Finally, this approach provides no information about the resolution of the image.

3.2 The Test-Function Approach

In this section, we present an optimization method to determine a weight function $W(\boldsymbol{x}, \boldsymbol{x}_s, \boldsymbol{x}_r)$. An advantage of this method is that it does not require the parametrization of the source-receiver geometry or the computation of (A.13). Moreover, it does not depend on the measured data itself, but only on the frequency band and the locations of sources and receivers. Once the weight function is determined with this method, the same weights can be applied to the inversion formula with any data set recorded with the same frequency band using the same source-receiver geometry.

Our test-function approach finds the weight function that is optimal in the sense that our point-spread function best approximates the Dirac delta function. To determine what "best" means, recall that the delta function has the property

$$f(\boldsymbol{x}) = \int \delta(\boldsymbol{x} - \boldsymbol{y}) f(\boldsymbol{y}) d\boldsymbol{y}, \tag{20}$$

for every smooth test function $f(\boldsymbol{x})$.

In this approach, we convolve a sequence of test functions with our point-spread function. If we can determine a weight function so that each test function can be recovered, that weight function should give a point-spread function close to a delta function. In particular, given any fixed image point $\boldsymbol{x}$, we choose non-overlapping real test functions $\chi_{jkl}(\boldsymbol{z})$ so that each $\chi_{jkl}(\boldsymbol{z})$ is a smooth approximation of the characteristic function supported on the $(j,k,l)^{\text{th}}$ voxel. Note that other test functions can be chosen and the results are not expected to be sensitive to the choice. First, we replace $\boldsymbol{x}-\boldsymbol{y}$ by $\boldsymbol{z}$ in the point-spread function $K_W(\boldsymbol{x}, \boldsymbol{y})$ and define

$$K_W(\boldsymbol{z}) = \frac{1}{(2\pi\bar{c})^3} \sum_{\boldsymbol{x}_s, \boldsymbol{x}_r} W(\boldsymbol{x}_s, \boldsymbol{x}_r) \times \int \omega^2 e^{-i\omega(\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}_s,\boldsymbol{x}) + \nabla_{\boldsymbol{x}}\tau(\boldsymbol{x},\boldsymbol{x}_r))\cdot\boldsymbol{z}} d\omega \tag{21}$$

where we drop the $\boldsymbol{x}$ dependence in W and K_W because each calculation is done with $\boldsymbol{x}$ fixed. We use the term "computational domain" to refer to the physical region of interest (denoted by the variable $\boldsymbol{y}$) which is the same as the image region (denoted by the variable $\boldsymbol{x}$). The point-spread function $K_W(\boldsymbol{z})$, however, depends on $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$; we call the computational region in $\boldsymbol{z}$ the "$\boldsymbol{z}$-computational domain". We want $K_W(\boldsymbol{z})$ to have the same property as delta function, and so require

$$\int K_W(\boldsymbol{z}) \chi_{jkl}(\boldsymbol{z}) d\boldsymbol{z} = \chi_{jkl}(\boldsymbol{0}). \tag{22}$$

Because the sources and receivers are discrete, the weight function W at each imaging point $\boldsymbol{x}$ is a vector whose dimension

equals the number of source-receiver pairs. Here we use the term *weights* to refer to the elements of this vector, and note that the weights are different for different $\boldsymbol{x}$.

To compute the left side of (22), where K_W is given by (21), we can use a 3-D *non-uniform fast Fourier transform* (NUFFT) (e.g., Greengard *et al.* 2004) to carry out the $\boldsymbol{z}$ integration, sample the result at the non-uniform locations $\omega(\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}_s,\boldsymbol{x}) + \nabla_{\boldsymbol{x}}\tau(\boldsymbol{x},\boldsymbol{x}_r))$, and then carry out the ω integration of (21). Alternatively, if we take the voxel size Δz less than a wavelength, we can approximate the left side of (22) via the mean value theorem:

$$\int\int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}(\tau(\boldsymbol{x}_s,\boldsymbol{x})+\tau(\boldsymbol{x},\boldsymbol{x}_r))\cdot\boldsymbol{z}} d\omega\chi_{jkl}(\boldsymbol{z})d\boldsymbol{z} \approx V\int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}(\tau(\boldsymbol{x}_s,\boldsymbol{x})+\tau(\boldsymbol{x},\boldsymbol{x}_r))\cdot\bar{\boldsymbol{z}}_{jkl}} d\omega. \quad (23)$$

Here $\bar{\boldsymbol{z}}_{jkl}$ is the center point of the $(j,k,l)^{\text{th}}$ voxel and V represents the volume of each voxel. Taking small voxels significantly decreases the computational cost of computing (21), at the expense of increasing the size and ill-conditioning of the system of equations (22).

Without loss of generality, we assume $\chi_{jkl}(\bar{\boldsymbol{z}}_{jkl}) = 1$. With (23), (22) becomes

$$\frac{1}{(2\pi\bar{c})^3}\sum_{\boldsymbol{x}_s,\boldsymbol{x}_r} W(\boldsymbol{x}_s,\boldsymbol{x}_r)\int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}(\tau(\boldsymbol{x}_s,\boldsymbol{x})+\tau(\boldsymbol{x},\boldsymbol{x}_r))\cdot\bar{\boldsymbol{z}}_{jkl}} d\omega = \begin{cases} 1/V & \text{if } \bar{\boldsymbol{z}}_{jkl}=\boldsymbol{0}, \\ 0 & \text{otherwise.}\end{cases} \quad (24)$$

The equations defined in (24) can be written as a linear system as follows. We enumerate the source-receiver pairs as $(\boldsymbol{x}_s^{(n)},\boldsymbol{x}_r^{(n)})$ for $n=1,\ldots,N$, and enumerate the voxels using the index $m=1,\ldots,M$. Thus, m is an enumeration of the indices j,k,l, and there are M test functions $\chi_m(\boldsymbol{z})$ for $m=1,\ldots,M$. The voxel centers are denoted $\bar{\boldsymbol{z}}_m$, and we assign the first voxel (pixel in 2-D) at the origin, i.e., $\bar{\boldsymbol{z}}_1 = 0$ (an example of a mesh in a 2-D $\boldsymbol{z}$-computational domain can be found in Fang (2008)). We multiply both sides of the mth equation by a positive constant λ_m. These normalization parameters do not change the equations, but they will play a role in the optimization scheme that we present later. In the numerical simulations, we will choose $\lambda_1 \geq 1$ and $\lambda_m = 1$ for $m \geq 2$ to emphasize the importance of the center peak of the point-spread function. Let $\boldsymbol{w}$ be the N dimensional weight vector whose elements are $W(\boldsymbol{x}_s^{(n)},\boldsymbol{x}_r^{(n)})$, and let $\boldsymbol{A}$ be the M-by-N matrix with $(m,n)^{\text{th}}$ entry A_{mn} defined to be

$$A_{mn} = \frac{\lambda_m}{(2\pi\bar{c})^3}\int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}(\tau(\boldsymbol{x}_s^{(n)},\boldsymbol{x})+\tau(\boldsymbol{x},\boldsymbol{x}_r^{(n)}))\cdot\bar{\boldsymbol{z}}_m} d\omega \quad (25)$$

Let $\boldsymbol{r}$ be the M-dimensional vector with components

$$r_m = \lambda_m\chi_m(\bar{\boldsymbol{z}}_m) = \begin{cases} \lambda_m/V & \text{if } \bar{\boldsymbol{z}}_m=\boldsymbol{0}, \\ 0 & \text{otherwise.}\end{cases} \quad (26)$$

The normalized version of (24) becomes the linear system $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{r}$, which we then solve to find the weight function $\boldsymbol{w}$. However, the matrix $\boldsymbol{A}$ is usually large, over-determined and ill-conditioned due to the fact that the number of unknowns (given by the number of source-receiver pairs) is usually much less than the number of knowns (given by the number of voxels or test functions), and redundancy is also introduced by our use of small voxels. Consequently, regularization is needed in order to obtain stable solutions. We use Tikhonov regularization (e.g., Aster *et al.* 2005) and define the functional $\mathcal{F}$

$$\mathcal{F}(\boldsymbol{w}) = \lambda_0^2\|\boldsymbol{w}\|_2^2 + \|\boldsymbol{A}\boldsymbol{w}-\boldsymbol{r}\|_2^2, \quad (27)$$

where $\|\cdot\|_2$ represents L^2 norm, and λ_0 is the Tikhonov regularization parameter or damper. Note that $\mathcal{F}(\boldsymbol{w})$ is always real. The optimization scheme we use to compute the weight function is to minimize $\mathcal{F}(\boldsymbol{w})$, i.e.,

$$\boldsymbol{w} = \text{argmin}\mathcal{F}(\boldsymbol{w}). \quad (28)$$

We can find an explicit formula for the solution of the regularized minimization problem from the *singular value decomposition* (SVD) method (Demmel 1997). Using $*$ for adjoint, we write the SVD of $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^*, \quad (29)$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are M-by-M and N-by-N unitary matrices, and $\boldsymbol{S}$ is a M-by-N rectangular diagonal matrix whose diagonal elements s_n are the singular values of $\boldsymbol{A}$. Thus the solution to (28) is

$$\boldsymbol{w} = \boldsymbol{V}\boldsymbol{Q}\boldsymbol{U}^*\boldsymbol{r}, \quad (30)$$

where

$$\boldsymbol{Q} = (\lambda_0^2\boldsymbol{I} + \boldsymbol{S}^*\boldsymbol{S})^{-1}\boldsymbol{S}^*. \quad (31)$$

The matrix $\boldsymbol{Q}$ is a N-by-M rectangular diagonal matrix with main diagonal entries

$$q_n = \frac{s_n}{s_n^2+\lambda_0^2}, \quad (32)$$

for $n=1,\ldots,N$. We call the weights calculated from this method the *optimization weights*. The optimization weights represent all the components of the vector of the *optimization weight function* W at a certain imaging point.

Solving (28) does not guarantee that the optimization weight function $\boldsymbol{w}$ is non-negative, by which we mean that both the real and imaginary parts of $\boldsymbol{w}$ are non-negative. For the corresponding MRI problem, Samsonov et al. (2003) and Bydder et al (2007) suspect that negative weights come from the instability of the inversion and result in worse reconstruction than non-negative weights. In our particular problem, the Tikhonov regularization parameter λ_0 controls the size of the optimization weights: a small λ_0 to results in wildly fluctuating weights, while a larger λ_0 will make most weights positive. An ideal choice of λ_0 would improve both the point-spread functions and the backprojection reconstruction. In the next section we discuss numerical results that provide some guidance on choosing λ_0.

To summarize, our algorithm for calculating the optimization weights as follows.

(i) List the source-receiver pairs $(\boldsymbol{x}_s^{(n)},\boldsymbol{x}_r^{(n)})$, where n is no greater than the number of source-receiver pairs.

(ii) Choose an imaging point $\boldsymbol{x}$.

(iii) Create a J-by-K-by-L mesh in the $\boldsymbol{z}$-computational domain. The mesh can be chosen the same for different imaging points. Write all the voxels in the mesh as an ordered list with the first voxel centered at the origin. The center point of mth voxel is $\bar{\boldsymbol{z}}_{\boldsymbol{m}}$, where $m \leq J\times K\times L$.

(iv) Set $\lambda_m = 1$ for $m \geq 2$, and choose λ_0 and λ_1. Both parameters should be positive.

(v) Construct the matrix $[A_{mn}]$ by (25) from mth test function and nth source-receiver pair.

(vi) Construct the vector $\boldsymbol{r}$ by (26).

(vii) Find the singular value decomposition of the matrix $\boldsymbol{A}$.

(viii) Construct the rectangular diagonal matrix $\boldsymbol{Q}$ by (32) from the singular values of $\boldsymbol{A}$ and λ_0.

(ix) Calculate the optimization weights $\boldsymbol{w}$ at $\boldsymbol{x}$ by (30).

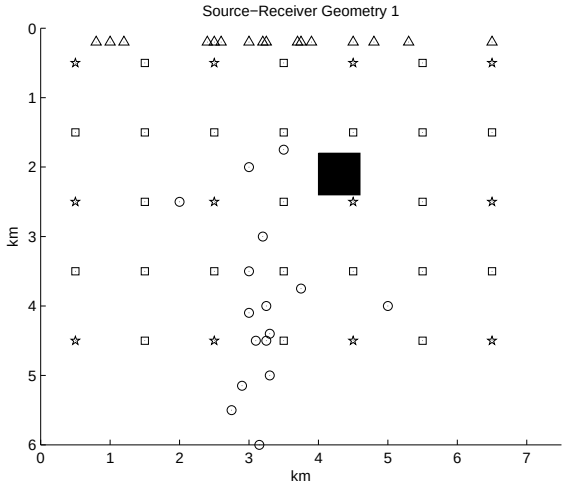


Figure 5. Geometry of sources (circles) and receivers (triangles). Black square locates the computational domain for reconstruction experiments done in section 3.4. Stars and squares indicate positions where point-spread functions are computed in section 3.5.1. Bisecting line segments are shown only at stars.

(x) Pick the next imaging point, go back to step (iii) and follow the same procedure until the optimization weights have been calculated for all the imaging points.

3.3 Numerical Simulations

We created an example with geometry typical of a passive seismic imaging problem (Fig. 5) to illustrate the test-function technique.

For the sake of simplicity, we consider only two-dimensional geometries (i.e., medium parameters $v(\boldsymbol{x})$, $c(\boldsymbol{x})$ and $\alpha(\boldsymbol{x})$ vary only in the x_1 and x_3 directions, and the sources and receivers are located in the plane $x_2 = 0$). Thus $\boldsymbol{x} = (x_1, x_3)$, $\boldsymbol{y} = (y_1, y_3)$, $\boldsymbol{z} = (z_1, z_3)$ and $\boldsymbol{k} = (k_1, k_3)$. We also assume the background wave speed is constant everywhere, i.e., $c(\boldsymbol{x}) = c = \bar{c}$; we take this constant to be 5km/s. In the constant-speed case, ray paths are straight lines and the calculation of the travel time, Green's function and $\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}, \boldsymbol{y})$ are simple evaluations of

$$\tau(\boldsymbol{x}, \boldsymbol{y}) = \frac{\|\boldsymbol{x} - \boldsymbol{y}\|}{c}, \tag{33}$$

$$G_0(\boldsymbol{x}, \boldsymbol{y}, \omega) = \frac{e^{i\frac{\omega}{c}\|\boldsymbol{x}-\boldsymbol{y}\|}}{4\pi\|\boldsymbol{x} - \boldsymbol{y}\|}, \tag{34}$$

$$\nabla_{\boldsymbol{x}}\tau(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x} - \boldsymbol{y}}{c\|\boldsymbol{x} - \boldsymbol{y}\|}. \tag{35}$$

The wave is emitted by each source and recorded by all receivers, and no two sources emit waves at the same time (so that waves from different sources do not interfere with each other).

We assume we start with real data and that the frequency ω satisfies $\omega \in \Omega = \{\omega | \omega_l \leq |\omega| \leq \omega_h\}$, where ω_h and ω_l are the highest and lowest positive frequencies, respectively. For applications that use real signals, such as seismic imaging, applying the Fourier transform to the time-domain measurements results in both positive and negative frequency content. In this case, $\boldsymbol{A}$ is a real matrix, because the imaginary part of the integrated function in equation (25) is an odd function. Thus, in our case, $\boldsymbol{A}$, $\boldsymbol{w}$ and $\boldsymbol{r}$ are all real quantities. However, $\boldsymbol{A}$, $\boldsymbol{w}$ and $\boldsymbol{r}$ might be complex for applications that use analytic signals, such as radar imaging, where recorded data is preprocessed by removing the negative frequency content.

Equation (23) does not change if the signs of $\boldsymbol{z}_m$ and ω are simultaneously reversed. This symmetry, together with the fact that we use both positive and negative frequencies, results in redundancies in the set of equations (24). We avoid these redundancies, and thus decrease the dimension of $\boldsymbol{A}$, by using only that half of the $\boldsymbol{z}$-computational domain with non-negative horizontal coordinates. Details regarding the calculation of $\boldsymbol{A}$ under these assumptions are discussed in Fang (2008).

In choosing the normalization parameters, we assume λ_1 corresponds to the test function supported on the pixel that contains the origin of the z-computational domain. If the number of test functions is large, λ_1 should be large. Otherwise, the solution to our least squares problem tends to zero as the number of equations in the problem becomes larger, since only the first entry of $\boldsymbol{r}$ is nonzero. This phenomenon can be seen from the QR factorization of $\boldsymbol{A}$. We choose $\lambda_1 \geq 1$, and $\lambda_m = 1$ for $m \geq 2$ so that point-spread functions at different locations have almost the same center height.

3.4 Regularization Experiments

The choice of regularization parameter or damper λ_0 affects both the resolution and stability of reconstructions. To evaluate the effects of λ_0, we created an example problem with a source-receiver configuration (Fig. 5) meant to mimic the locally recorded earthquake imaging environment with irregular spacing of sources and receivers [locations are based on a microearthquake network from central California (e.g., Roecker *et al.* 2006)].

The objective is to reconstruct a point-like scatterer 10 m in extent located at $\boldsymbol{y}_0 = (4.3, 2.1)$. We assume a frequency band from 5Hz to 50 Hz and -5 Hz to -50 Hz; consequently the smallest wavelength is 100m. We take the normalization parameter λ_1 to be 20, and set λ_0 to 1, 500, and finally 10,000.

We show plots of both the point-spread function and reconstructions. For the reconstructions, the computational domain is the square $\|\boldsymbol{x} - (4.3, 2.1)\|_\infty \leq 0.6$ km (the black square shown in Fig. 5), and the mesh size is 10 m. Hence we have 1200×1200 imaging points in the computational domain for the reconstructions.

For the plots of the point-spread function, the $\boldsymbol{z}$-computational domain is $\|\boldsymbol{z}\|_\infty \leq 0.6$ km, again with mesh size 10 m. However, because of the symmetry of (23) mentioned above, we carry out computations only for the half of the region with non-negative horizonal coordinates (i.e., 600×1200 points), and consequently we have only 600×1200 test functions. Because the mesh size is smaller than the smallest wavelength, we are able to use the approximation (23).

Mathematically, the scatterer is defined as

$$\alpha(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \|\boldsymbol{y} - \boldsymbol{y}_0\|_\infty \leq 5\text{m}, \\ 0 & \text{otherwise}, \end{cases} \tag{36}$$

where $\|\cdot\|_\infty$ represents the infinity norm and is defined to be the maximum absolute value of all elements of a vector. Note that the scatterer size is less than the shortest wavelength (100m) in our simulation.

The simulation data are calculated using equation (7) with

$\alpha(\boldsymbol{x})$ defined by (36). As the background speed is constant,

$$a(\boldsymbol{x}_s, \boldsymbol{y}, \boldsymbol{x}_r) = \frac{1}{16\pi^2 \|\boldsymbol{x}_s - \boldsymbol{y}\| \|\boldsymbol{y} - \boldsymbol{x}_r\|}, \tag{37}$$

$$\phi(\boldsymbol{x}_s, \boldsymbol{y}, \boldsymbol{x}_r) = \frac{\|\boldsymbol{x}_s - \boldsymbol{y}\| + \|\boldsymbol{y} - \boldsymbol{x}_r\|}{c}. \tag{38}$$

Since the size of the scatterer is relatively small, we use the following approximation of the scattered wave:

$$\begin{aligned} u_D^s(\boldsymbol{x}_r, \boldsymbol{x}_s, \omega) &= \frac{\omega^2}{c^2} \int a(\boldsymbol{x}_s, \boldsymbol{y}, \boldsymbol{x}_r) e^{i\omega\phi(\boldsymbol{x}_s, \boldsymbol{y}, \boldsymbol{x}_r)} \alpha(\boldsymbol{y}) d\boldsymbol{y}, \\ &\approx 10^{-4} \frac{\omega^2}{c^2} a(\boldsymbol{x}_s, \boldsymbol{y}_0, \boldsymbol{x}_r) e^{i\omega\phi(\boldsymbol{x}_s, \boldsymbol{y}_0, \boldsymbol{x}_r)}, \end{aligned} \tag{39}$$

where 10^{-4} comes from the area of the point-like scatterer. Note that this approximation is equivalent to setting

$$\alpha(\boldsymbol{y}) = 10^{-4} \delta(\boldsymbol{y} - \boldsymbol{y}_0), \tag{40}$$

which justifies calling it a point-like scatterer.

To simulate band-limited data, we set $u_D^s(\boldsymbol{x}_r, \boldsymbol{x}_s, \omega)$ to zero outside $\Omega = \{\omega | \omega_l \leq |\omega| \leq \omega_h\}$, and take the inverse Fourier transform of $u_D^s(\boldsymbol{x}_r, \boldsymbol{x}_s, \omega)$ to recover the time-domain scattered wave $\hat{u}_D^s(\boldsymbol{x}_r, \boldsymbol{x}_s, t)$:

$$\hat{u}_D^s(\boldsymbol{x}_r, \boldsymbol{x}_s, t) = \int_\Omega e^{-i\omega t} u_D^s(\boldsymbol{x}_r, \boldsymbol{x}_s, \omega) d\omega, \tag{41}$$

$$= \frac{a(\boldsymbol{x}_s, \boldsymbol{y}_0, \boldsymbol{x}_r)}{10^4 c^2} \int_\Omega \omega^2 e^{-i\omega(t - \phi(\boldsymbol{x}_s, \boldsymbol{y}_0, \boldsymbol{x}_r))} d\omega. \tag{42}$$

Because we do not have zero frequency data, backprojection by (A.16) cannot recover the actual image even when the sources and receivers are equally spaced. Nevertheless, the correct locations of discontinuities can be recovered (e.g., Bleistein *et al.* 2001), and more information about a point scatterer can be recovered if the image is scaled properly. If we use $\alpha(\boldsymbol{y}) = R\delta(\boldsymbol{y} - \boldsymbol{y}_0)$ in (10), the image value at the scatterer $\boldsymbol{y}_0$ is

$$\begin{aligned} I_W(\boldsymbol{y}_0) &= \frac{1}{(2\pi\bar{c})^3} \sum_{\boldsymbol{\sigma}} W(\boldsymbol{y}_0, \boldsymbol{\sigma}) \int_\Omega \frac{c^2(\boldsymbol{y}_0)}{a(\boldsymbol{y}_0, \boldsymbol{\sigma})} \omega^2 \\ &\quad \times \int_{R^3} \frac{R\delta(\boldsymbol{y} - \boldsymbol{y}_0)}{c^2(\boldsymbol{y})} a(\boldsymbol{y}, \boldsymbol{\sigma}) e^{-i\omega(\phi(\boldsymbol{y}_0, \boldsymbol{\sigma}) - \phi(\boldsymbol{y}, \boldsymbol{\sigma}))} d\boldsymbol{y} d\omega \end{aligned} \tag{43}$$

$$= \frac{R}{(2\pi\bar{c})^3} \sum_{\boldsymbol{\sigma}} W(\boldsymbol{y}_0, \boldsymbol{\sigma}) \int_\Omega \omega^2 d\omega. \tag{44}$$

This suggests that the image value is R multiplied by the k-space "volume" $V_K(\boldsymbol{y}_0)$, which is given by

$$\begin{aligned} V_K(\boldsymbol{x}) &= \frac{1}{(2\pi)^3} \int d\boldsymbol{k} \\ &\approx \frac{1}{(2\pi\bar{c})^3} \sum_{\boldsymbol{\sigma}} W(\boldsymbol{x}, \boldsymbol{\sigma}) \int_\Omega \omega^2 d\omega. \end{aligned} \tag{45}$$

[To compare with the case of a continuum of sources and receivers, see equations (A.17) and (A.18).] If we scale the image by dividing it by the k-space "volume" (45), then we can recover the amplitude R at the scatterer. The quantity R is exactly the center peak of the point-spread function at $\boldsymbol{x}$. Since we choose λ_1 large enough so that the point-spread function at different locations in our computational domain has almost the same peak value, the scaling factor $V_K(\boldsymbol{x})$ is almost constant everywhere. Note that scaling the point-spread function changes neither the resolution nor the stability of backprojection. Because our purpose here is to examine resolution and stability, we ignore the scaling in our simulations.

Note that in all our simulations, the units on the horizonal and vertical axes in the point-spread function plot, in the reconstructed image, and in the half-maximum-value contour plot are km. The units of point-spread function values are $1/km^2$. The image $I(x)$ is dimensionless in our formulation. The units on the horizonal and the vertical axes in the plots of bisecting line segments are $1/km$. Because of our choice of scaling, the weight function $\boldsymbol{w}$ is dimensionless. Throughout the sections 3.4 and 3.5, we consider these units as default units, unless explicitly stated otherwise.

3.4.1 Simulation Results

The results shown in columns *2* through *4* in Fig. 6 illustrate how the regularization parameter λ_0 affects reconstruction when optimization weights are used. For comparison purposes, column *1* shows results from assuming constant weights (i.e., $\boldsymbol{w} = 1$) scaled so that the point-spread function has the same height as in the optimization weights case. Row *a* in Fig. 6 shows the bisecting line segments defined by (19) shaded according to the corresponding weight. We plot only the positive frequency part; the corresponding negative-frequency bisecting line segment simply has the opposite sign. The constant weights case shown in column 1 is simply a plot of all the bisecting line segments with the same grey scale value. Columns 2-4 show the case of optimization weights with increasing choices of λ_0. The distribution of bisecting line segments is the same as for the constant weights, but many of the bisecting line segments have small weights and cannot be seen with this color scale. This is particularly true for $\lambda_0 = 1$, where weight values lie in a wide range, roughly from -200 to 300. As expected, the size and range of the weight values decreases as damping increases. In the case of $\lambda_0 = 10,000$ (column 4), the weight values are small and most are positive. One phenomenon we see in Fig. 6(a4) is that the directions with fewer bisectors have higher weight values and the directions with more bisectors have smaller weight values.

Row *b* shows the point-spread function at the scatterer $\boldsymbol{y}_0$. The point-spread function computed with constant weights (column 1) has a high peak at the center, but is elongated in the vertical direction, meaning that the reconstruction at this point has better resolution in the horizontal direction. There are many more bisecting line segments pointing in the horizontal direction, and weighting each direction equally degrades the resolution in the vertical direction. By contrast, the point-spread functions calculated using the optimization weights shown in columns 2-4 are much narrower in the vertical direction than the one calculated using constant weights. The similarity in the shape of the function for different choices of λ_0 indicate that resolution is not overly sensitive to the choice of damper.

Row *c* shows reconstructions of the scatterer using simulation data without noise. In most cases, the reconstruction is similar to the point-spread function shown in row *b*. Note that the point-spread function is calculated using (18) and the image is calculated using (10). They are related by equation (16) and should be almost the same if $\alpha(\boldsymbol{x})$ is a point-like scatterer. However, a difference between them arises as we pass from (11) to (15), namely that we have made the approximations (12) and (13), which are good when $\boldsymbol{x}$ is close to $\boldsymbol{y}$, but not when they are far apart. The reconstruction simulation result suggests these approximations work well with constant weights, but require regularization when using optimization weights. In particular, a choice of $\lambda_0 = 1$ shows a localized scatterer at the center, but the reconstruction is corrupted away from the scatterer, because the errors made in using the approximations (12) and (13) are significantly amplified by large magnitude weights.

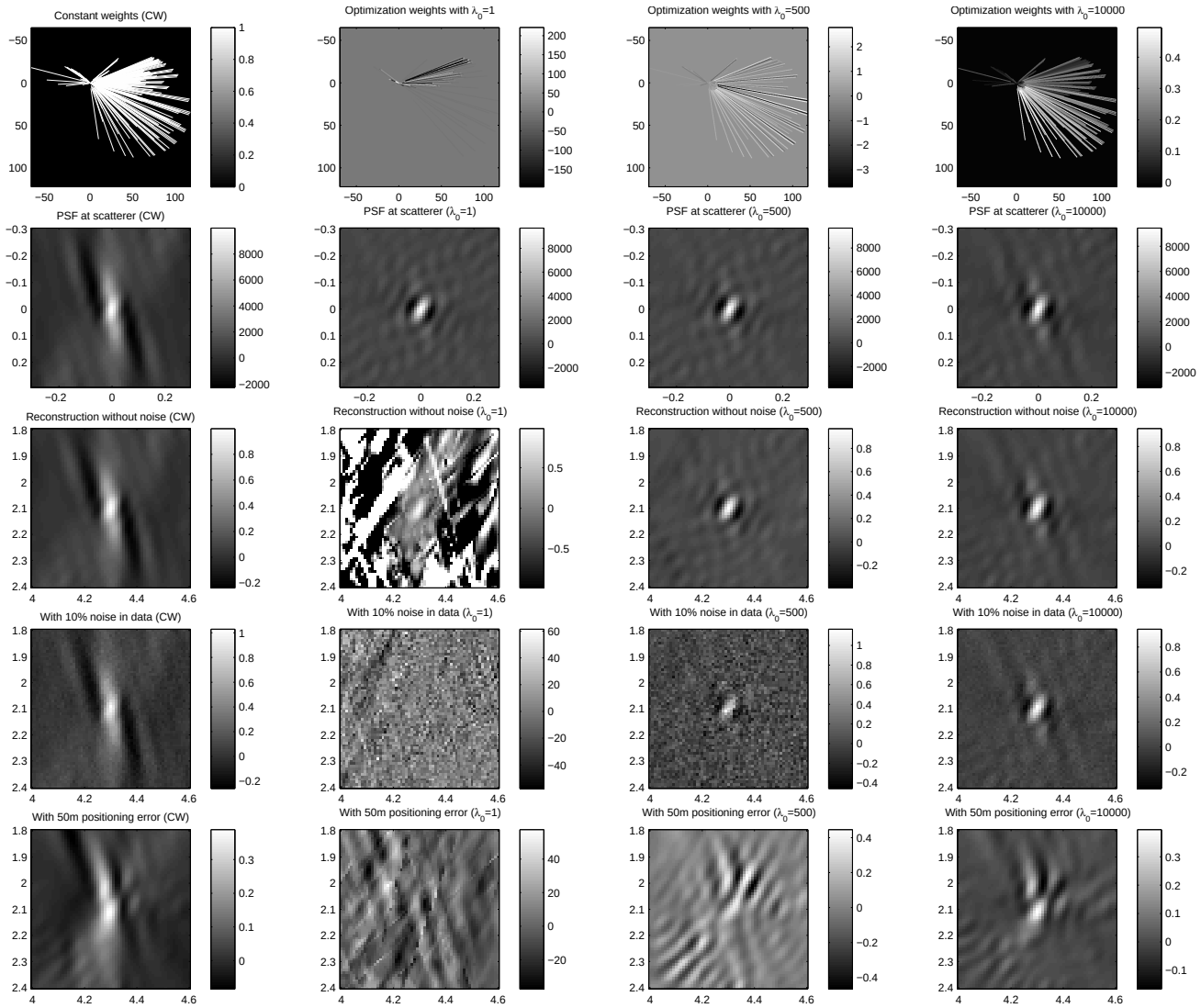


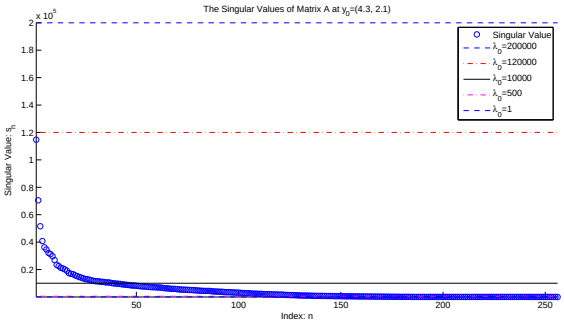
Figure 6. Bisecting line segments (first row), point-spread functions (second row), and reconstructions (three lower rows) for constant weight (column 1), $\lambda_0 = 1$ (column 2), $\lambda_0 = 500$ (column 3), and $\lambda_0 = 10,000$ (column 4) using source-receiver configuration in the geometry shown in Fig. 5. The reconstructions were done using data without noise (third row), with 10% Gaussian noise (fourth row), and with up-to-50-meter positioning error in the source-receiver locations (fifth row). The computational domain in these reconstructions corresponds to the black square in the geometry in Fig. 5. For the figures on the first row, the horizontal axis is k_1, and the vertical axis is k_3. The units on both axes are $1/\text{km}$. For the figures on the second row, the horizontal axis is z_1, and the vertical axis is z_3. The units on both axes are km. For the figures on the third to fifth row, the horizontal axis is x_1, and the vertical axis is x_3. The units on both axes are km.

The results in columns 3 and 4 show that these artifacts can be eliminated by proper damping, and we can achieve a significantly better reconstruction than when using constant weights.

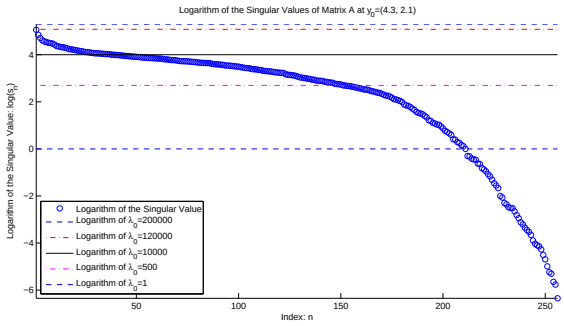
Row *d* shows the reconstruction with 10% Gaussian noise added to the simulation data. Although the reconstruction in each case is noisier, and fails in the case of small damping (column 2), in general both the constant weights and properly damped optimized weights are reasonably insensitive to noise at this level, although again the reconstruction using properly regularized optimized weights is superior.

Row *e* shows the reconstruction with random errors of up to 50 meters added to the locations of the sources and receivers (50 meters is the half minimum wavelength in this case). We note that errors in the earthquake source location can easily give rise to 50-meter positioning errors; for other applications such as radar, positioning errors of half a wavelength are difficult to avoid. For positioning errors larger than a wavelength, other methods such as autofocus techniques (e.g., Jakowatz *et al.* 1996) would need to be applied to compensate for this type of error. Again, proper damping plays a significant role as the two cases with small damping (columns 2 and 3) fail to locate the scatterer in the correct position (the peak in the $\lambda_0 = 500$ example is about 140 meters away from the true location). Both the constant weight and large damping results show a centralized peak in the correct position, with some smaller amplitude side lobes, but that obtained from optimization weights is more strongly localized.

As shown in (32), the main effect of the damper is to lessen the influence of small singular values. An examination of the singular values for $\boldsymbol{A}$ (Figures 7(a) and 7(b)) shows that our preferred value of $\lambda_0 = 10,000$ is greater than 90% of the singular



(a) Singular values of matrix $\boldsymbol{A}$ at $\boldsymbol{y}_0$. The horizontal lines correspond to the regularization parameters $\lambda_0 = 1$, $\lambda_0 = 500$, $\lambda_0 = 10,000$, $\lambda_0 = 120,000$ and $\lambda_0 = 200,000$. The $\lambda_0 = 1$ line and the $\lambda_0 = 500$ line are indistinguishable from the horizontal axis.



(b) Logarithm of the singular values of matrix $\boldsymbol{A}$ at $\boldsymbol{y}_0$. Meaning of symbols same as above.

Figure 7.

values. This suggests that a relatively small number of bisectors, mostly those from source-receiver combinations whose bisector directions are relatively sparse, are significantly affected by the optimization. Optimization weights from regularization parameters that exceed the values of the largest singular value ($\lambda_0 = 120,000$ and $\lambda_0 = 200,000$) are all positive and nearly constant. The resulting point-spread functions and reconstructions are not very different from the results with constant weights. Because λ_0 controls the magnitude of our weight vector $\boldsymbol{w}$, the term $\|\boldsymbol{w}\|_2^2$ becomes dominant in the minimization scheme when λ_0 is chosen this large.

The results of this test show that the choice of λ_0 depends on the noise level in the data and the accuracy of the locations of sources and receivers. There is no simple formula for choosing λ_0, but we see that as λ_0 is increased, more of the optimization weights tend to be positive. This suggests that we choose λ_0, so that most of the optimization weight values are positive. In this paper, the simulation results we show in the next section were calculated with values of λ_0 chosen using this criterion.

3.5 Numerical Examples of Weights in Seismic Imaging

We carried out some numerical experiments to assess the improvement of resolution at selected imaging points due to the use of the weights obtained by our method. We show the results of one experiment and simply summarize the results of others, which are reported more fully in Fang (2008).

Table 1. List of parameters. The reconstructions used 16 sources and 16 receivers, and the wave speed c was 5km/s. f is frequency.

$\lvert f\rvert$	$\boldsymbol{z}$-computational domain	Δz	λ_0	λ_1
5Hz-50Hz	$\lVert\boldsymbol{z}\rVert_\infty \leq 0.6$km	0.01km	10^4	20

3.5.1 *Local/Regional Passive Source*

We applied our test-function approach to the source-receiver geometry shown in Fig. 5. We calculate the point-spread functions at 7×5 locations (indicated by stars and squares shown in Fig. 5) where adjacent imaging points are separated by 1 km in both horizontal and vertical directions.

Comparing the point-spread functions computed with constant weights (Fig. 8) and with optimization weights (Fig. 9), we find that the point-spread functions with the optimization weights are sharper and more focused than the ones calculated with constant weights, especially for the imaging points close to the center of our computational domain.

We quantify the resolution at a point $\boldsymbol{x}_0$ by the volume (or area in 2-D) of the enclosed region determined by the half-maximum-value contour of the point-spread function. This contour also provides not only a quantitative measure of the improvement but also a rough estimate of resolution in every direction. For example, for the imaging point shown at the center, the half-maximum-value area for constant weights is 15102 m^2, while that from the optimization weights is 1649 m^2. At this point, the half-maximum-value contour from optimization weights has an elliptical shape. The best resolution, which is in the direction of the minor axis, is about 30 m , while the worst resolution, in the direction of the major axis, is about 70 m . As the minimum wavelength in this example is 100 m, we obtain sub-wavelength resolution that is roughly consistent with the resolution limit of one fourth the minimum wavelength (e.g., Bleistein *et al.* 2001).

An understanding of the importance of the weights can be obtained by simply plotting the bisecting line segments. Fig. 11 shows a plot of the bisecting line segments colored according to the optimization weights. We see that in general, bisectors in areas of lower density are assigned higher weights, while the redundant information from regions of high density are assigned lower weights.

3.5.2 *Summary of Other Experiments*

We also tested our approach with regular source-receiver geometries such as active source profile and cross-hole. In these cases the bisecting line segments are more evenly distributed, the constant weights work relatively well, and optimization weights don't improve resolution much. In such cases, computing optimization weights may not be worth the extra effort.

We also tested this approach with the same geometry but different frequency content. The point-spread functions from different frequency content roughly have the same shape, but very different scale. To explain this phenomena, we consider the region covered by the set of bisecting line segments. This region is a discrete approximation to the data-collection manifold described in the appendix [below (A.18)]. For a narrower and lower frequency band, the region covered by the set of bisecting line segments is a relatively small area, while for the higher and broader frequency band, this region is larger. The larger region results in a better point-

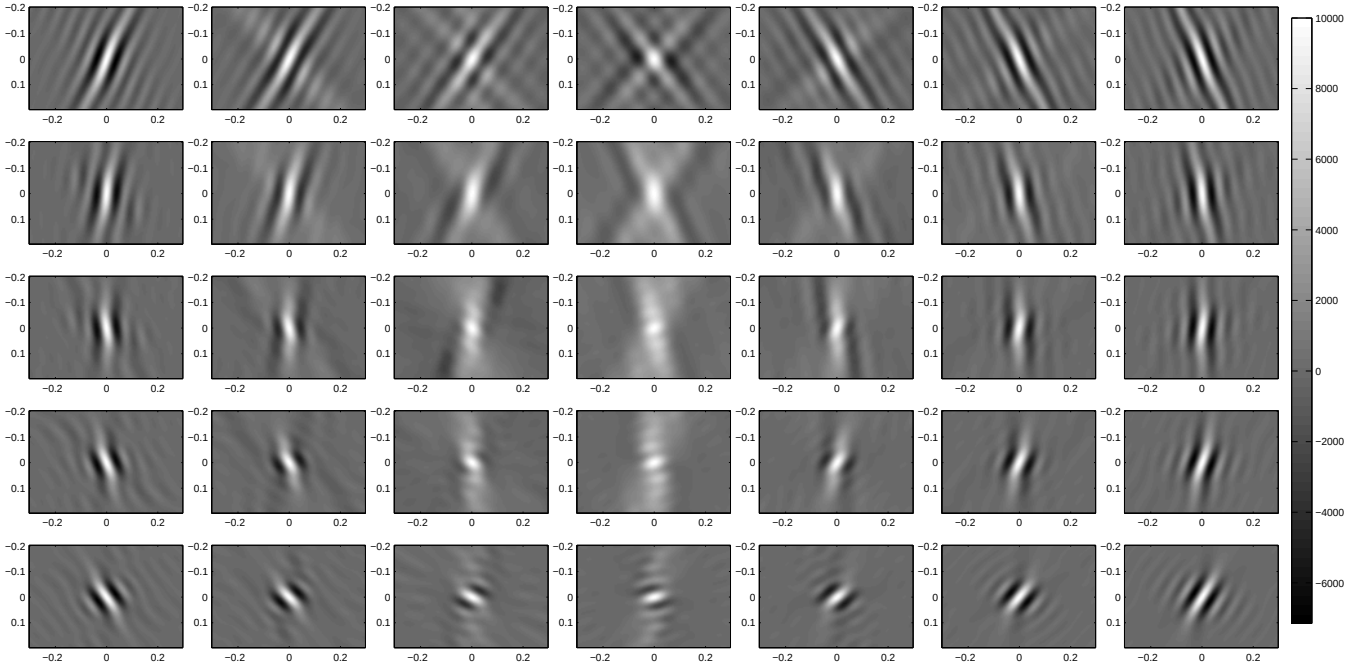


Figure 8. Plots of the point-spread functions using constant weights. The horizontal axis is z_1 and the vertical axis is z_3. The units on each axis are km. The units of the point-spread function values are $1/\text{km}^2$. The locations correspond to the 7×5 different imaging points designated by the stars and squares in the geometry in Fig. 5.

spread function. Consequently, to produce an image with better resolution, a broader frequency band should be used.

4 CONCLUSIONS

We have developed a method to handle imaging problems in which the sensor geometry is irregular and sparse. Our analysis shows that a plot of the bisecting line segments can be used as a guide in the reconstruction process. If the bisecting-line-segment plot is regular, then good images can be formed by simply adding data from the various source-receiver pairs, and resolution in various directions can be estimated from the plot. On the other hand, if the plot of bisecting line segments is irregular and has gaps, then data from different source-receiver pairs should be weighted differently. In this case, our method produces weights that gives rise to the "best" point-spread function; once the weights are obtained, they can then used to weight the data properly to produce the "best" image. In other words, this approach provides not only an image, but also a point-spread function from which image resolution can be determined.

In calculating the weights, we found that in order to gain stability with respect to noise and sensor-positioning errors, the regularization parameter should be chosen so that the optimization weight values are mostly positive.

Our method can be used as a tool for planning an experiment: given sensor positions and bandwidths, we can plot the corresponding point-spread function and thus predict resolution in different directions at desired locations. For example, our tests with sensors of different bandwidths show that a broad frequency band is needed to reconstruct small objects.

APPENDIX A: BACKPROJECTION RECONSTRUCTION

In this section, we briefly review a derivation (e.g., Bleistein *et al.* 2001; Beylkin 1985) of a formula for inverting (7) in the ideal case where the data is known on a three dimensional set of points $(\omega, \boldsymbol{x}_s, \boldsymbol{x}_r)$. We parameterize the source-receiver geometry by the two dimensional variable $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$, and use $\boldsymbol{\sigma}$ to represent $(\boldsymbol{x}_s, \boldsymbol{x}_r)$.

In the case of 3-D densely sampled data, we look for an inverse to (7) in terms of a filtered adjoint:

$$I(\boldsymbol{x}) = \int\int b(\boldsymbol{x}, \boldsymbol{\sigma}) u_D^s(\boldsymbol{\sigma}, \omega) e^{-i\omega\phi(\boldsymbol{x},\boldsymbol{\sigma})} d\omega d\boldsymbol{\sigma}. \qquad \text{(A.1)}$$

Here $I(\boldsymbol{x})$ is called the image of $\alpha(\boldsymbol{x})$ and the expression for $b(\boldsymbol{x}, \boldsymbol{\sigma})$ is determined below.

We can carry out the ω integration in equation (A.1) to obtain

$$I(\boldsymbol{x}) = \int b(\boldsymbol{x}, \boldsymbol{\sigma}) \hat{u}_D^s(\boldsymbol{\sigma}, \phi(\boldsymbol{x}, \boldsymbol{\sigma})) d\boldsymbol{\sigma}. \qquad \text{(A.2)}$$

This equation says the data from the source-receiver pair $\boldsymbol{\sigma}$ at time t is first filtered by $b(\boldsymbol{x}, \boldsymbol{\sigma})$, then backprojected to (i.e., spread out over) the surface $\{\boldsymbol{x} | \phi(\boldsymbol{x}, \boldsymbol{\sigma}) = t\}$. The image $I(\boldsymbol{x})$ is then the superposition of all the backprojected data from all source-receiver pairs. For this reason, this method is often referred to as the backprojection method.

To determine $b(\boldsymbol{x}, \boldsymbol{\sigma})$, we analyze the relationship between the image $I(\boldsymbol{x})$ and the actual perturbation $\alpha(\boldsymbol{x})$. To do this, we plug (7) into (A.1) and change the order of integration. This results in

$$I(\boldsymbol{x}) = \int K(\boldsymbol{x}, \boldsymbol{y}) \alpha(\boldsymbol{y}) d\boldsymbol{y}, \qquad \text{(A.3)}$$

where $K(\boldsymbol{x}, \boldsymbol{y})$ is the *point-spread function*

$$K(\boldsymbol{x}, \boldsymbol{y}) = \int \frac{\omega^2}{c^2(\boldsymbol{y})} a(\boldsymbol{y}, \boldsymbol{\sigma}) b(\boldsymbol{x}, \boldsymbol{\sigma}) e^{-i\omega(\phi(\boldsymbol{x},\boldsymbol{\sigma}) - \phi(\boldsymbol{y},\boldsymbol{\sigma}))} d\omega d\boldsymbol{\sigma}. \qquad \text{(A.4)}$$

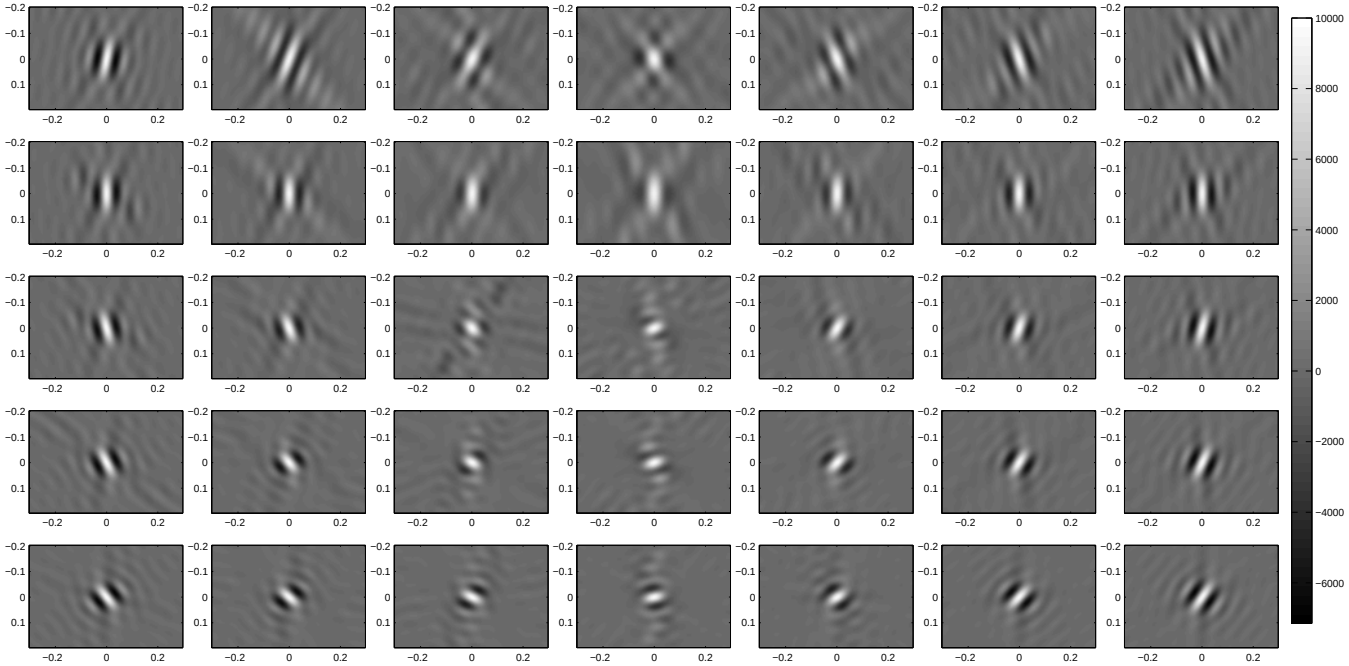


Figure 9. Plots of point-spread functions using optimization weights. The units and symbols are the same as in Fig. 8.

To obtain an optimal image, we would like the right hand side of the equation (A.3) to be a convolution of $\alpha(\boldsymbol{x})$ with $\delta(\boldsymbol{x})$, which we write as

$$\delta(\boldsymbol{x}-\boldsymbol{y}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})}d\boldsymbol{k}. \tag{A.5}$$

We analyze (A.4) by the method of stationary phase (e.g., Duistermaat 1995; Grigis *et al.* 1994). The main contributions to (A.4) come from the critical points, which are obtained from differentiating the phase of (A.4) with respect to ω and $\boldsymbol{\sigma}$. We find that the critical conditions are

$$\phi(\boldsymbol{x},\boldsymbol{\sigma}) = \phi(\boldsymbol{y},\boldsymbol{\sigma}), \tag{A.6}$$

$$\omega\nabla_{\boldsymbol{\sigma}}\phi(\boldsymbol{x},\boldsymbol{\sigma}) = \omega\nabla_{\boldsymbol{\sigma}}\phi(\boldsymbol{y},\boldsymbol{\sigma}). \tag{A.7}$$

We assume here that there are solutions to (A.6) only when $\boldsymbol{x}=\boldsymbol{y}$ (e.g., Bleistein *et al.* 2001; Beylkin 1985). In the neighborhood of $\boldsymbol{x}=\boldsymbol{y}$, we can use the following approximations in (A.4).

$$b(\boldsymbol{x},\boldsymbol{\sigma}) \approx b(\boldsymbol{y},\boldsymbol{\sigma}). \tag{A.8}$$

$$\omega(\phi(\boldsymbol{x},\boldsymbol{\sigma})-\phi(\boldsymbol{y},\boldsymbol{\sigma})) \approx \omega\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\cdot(\boldsymbol{x}-\boldsymbol{y}). \tag{A.9}$$

In (A.4) we replace $b(\boldsymbol{x},\boldsymbol{\sigma})$ by $b(\boldsymbol{y},\boldsymbol{\sigma})$ and make the change of variables

$$(\omega,\boldsymbol{\sigma}) \mapsto \boldsymbol{k} = \omega\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma}) \tag{A.10}$$

to get

$$K(\boldsymbol{x},\boldsymbol{y}) \approx \int \frac{\omega^2(\boldsymbol{k})}{c^2(\boldsymbol{y})}a(\boldsymbol{y},\boldsymbol{\sigma}(\boldsymbol{k}))b(\boldsymbol{y},\boldsymbol{\sigma}(\boldsymbol{k}))\,|J(\boldsymbol{k})|\,e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})}d\boldsymbol{k} \tag{A.11}$$

where $J(\boldsymbol{k}) = \det(\partial(\omega,\boldsymbol{\sigma})/\partial\boldsymbol{k})$. Note that

$$\frac{1}{J(\boldsymbol{k})} = \det\left(\frac{\partial\boldsymbol{k}}{\partial(\omega,\boldsymbol{\sigma})}\right), \tag{A.12}$$

$$= \omega^2\det\begin{pmatrix}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\\ \frac{\partial}{\partial\sigma_1}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\\ \frac{\partial}{\partial\sigma_2}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\end{pmatrix}, \tag{A.13}$$

$$= \omega^2 h(\boldsymbol{x},\boldsymbol{\sigma}). \tag{A.14}$$

Comparing (A.11) with (A.5), we see that we should choose $b(\boldsymbol{y},\boldsymbol{\sigma}(\boldsymbol{k}))$ such that $\omega^2(\boldsymbol{k})c^{-2}(\boldsymbol{y})a(\boldsymbol{y},\sigma(\boldsymbol{k}))b(\boldsymbol{y},\boldsymbol{\sigma}(\boldsymbol{k}))\,|J(\boldsymbol{k})| = (2\pi)^{-3}$. Thus we have an explicit form for $b(\boldsymbol{y},\boldsymbol{\sigma})$, namely

$$b(\boldsymbol{y},\boldsymbol{\sigma}) = \frac{c^2(\boldsymbol{y})\,|h(\boldsymbol{y},\boldsymbol{\sigma})|}{(2\pi)^3 a(\boldsymbol{y},\boldsymbol{\sigma})}. \tag{A.15}$$

Using the above expression for $b(\boldsymbol{y},\boldsymbol{\sigma})$ in equation (A.1), we obtain the inverse formula to reconstruct $\alpha(\boldsymbol{x})$, namely,

$$I(\boldsymbol{x}) = \int\int \frac{c^2(\boldsymbol{x})\,|h(\boldsymbol{x},\boldsymbol{\sigma})|}{(2\pi)^3 a(\boldsymbol{x},\boldsymbol{\sigma})}u_D^s(\boldsymbol{\sigma},\omega)e^{-i\omega\phi(\boldsymbol{x},\boldsymbol{\sigma})}d\omega d\boldsymbol{\sigma}. \tag{A.16}$$

Because our data are band-limited and the source-receiver geometry is finite, the integration region of (A.11) is a bounded domain $\Omega_{\boldsymbol{x}}$ rather than all of $\mathbb{R}^3$ space. As a result, $\alpha(\boldsymbol{x})$ cannot be perfectly recovered. However, the point-spread function $K(\boldsymbol{x},\boldsymbol{y})$ quantifies the degree to which our image faithfully represents the true $\alpha(\boldsymbol{x})$. Using (A.15) in equation (A.11), we get

$$K(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{(2\pi)^3}\int |h(\boldsymbol{x},\boldsymbol{\sigma})| \int \omega^2 e^{-i\omega\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x},\boldsymbol{\sigma})\cdot(\boldsymbol{x}-\boldsymbol{y})}d\omega d\boldsymbol{\sigma}. \tag{A.17}$$

The degree to which $K(\boldsymbol{x},\boldsymbol{y})$ approximates a delta function is determined by the source-receiver geometry and the frequency band of the data. In the continuum case, we can make change of variables

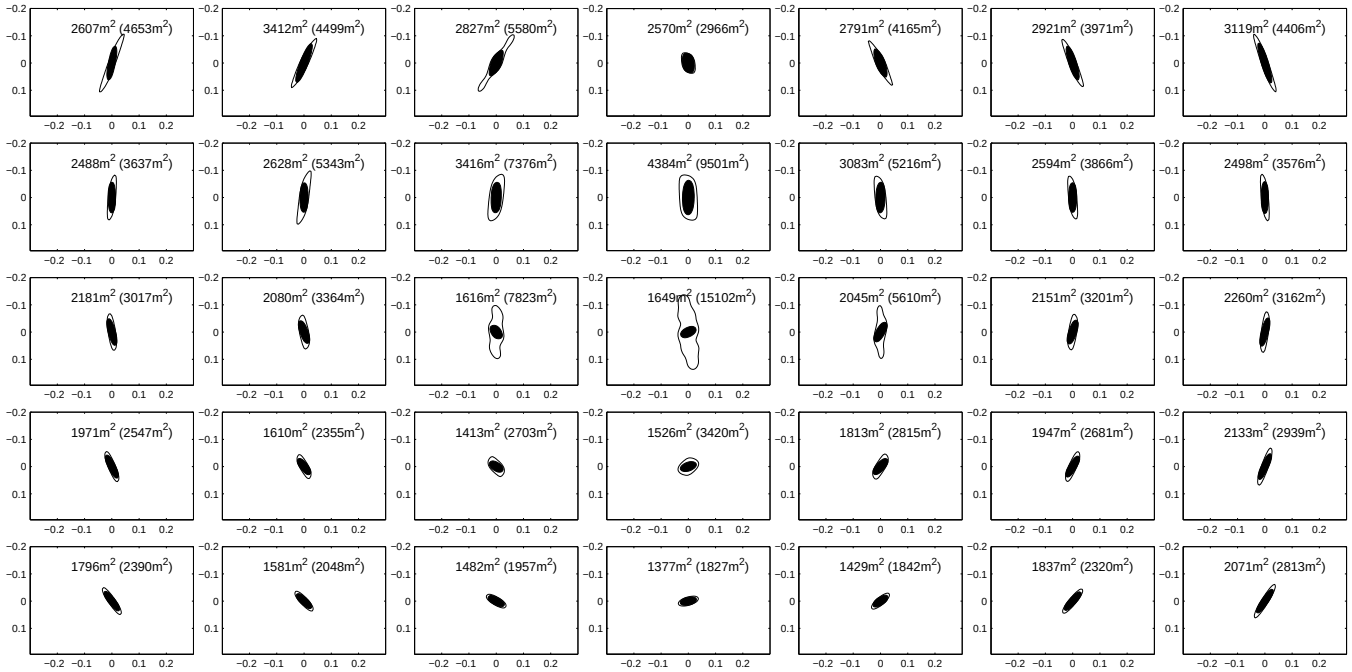


Figure 10. Half-maximum-value contours of the point-spread functions calculated with optimization weights (the contours filled with black) and constant weights (the outer contours). The horizontal axis is z_1 and the vertical axis is z_3. The units on each axis are km. The 7×5 different imaging points correspond to both stars and squares in the geometry in Fig. 5. The number shown outside the parenthesis is half-maximum-value area with optimization weights. The number shown inside the parenthesis is half-maximum-value area with constant weights.

(A.10) and define $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$ in (A.17). Thus we obtain

$$K(\boldsymbol{z}) = \frac{1}{(2\pi)^3} \int_{\Omega_{\boldsymbol{x}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{z}} d\boldsymbol{k}, \qquad \text{(A.18)}$$

where the data-collection manifold $\Omega_{\boldsymbol{x}}$ is the set in $\boldsymbol{k}$-space, obtained from the change of variables (A.10), that corresponds to the bandwidth (set of frequencies ω) and survey geometry (set of bisectors $\nabla\phi(\boldsymbol{x}, \boldsymbol{\sigma})$ as $\boldsymbol{\sigma}$ ranges over the sources and receivers). If the data-collection manifold $\Omega_{\boldsymbol{x}}$ is the whole space $\mathbb{R}^3$, then (A.18) is a delta function, which justifies our choice for $b(\boldsymbol{y}, \boldsymbol{\sigma})$.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant number CMG-04000000 and contract number DMS-0327634 and by the Air Force Office of Scientific Research* under agreement number FA9550-06-1-0017. Some computing resources were provided by the DARPA/Geo* project at RPI. Other computing resources were provided by the Inverse Problems Center at Rensselaer.

We would like to acknowledge Joyce Mclaughlin, Daniel Renzi, David Isaacson, Randolph Franklin, Kui Lin, Jing Hu, Ashley Thomas, Fengyan Li, Leslie Greengard, Il-Young Son, Trond Varslot, Ling Wang, Ning Zhang, Polina Zheglova, Hector Morales, Matthew Ferrara, Sava Dediu, and Jeong-Rock Yoon for useful discussions and suggestions.

* Consequently the U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

REFERENCES

R.C. Aster, B. Borchers and C. Thurber, *Parameter Estimation and Inverse Problems*, Academic Press (2005) p. 301.

Aurenhammer, F., *Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure*, ACM Computing Surveys, 23(3) (1991), pp. 345-405.

Backus, G. & Gilbert, F., *The Resolving Power of Gross Earth Data*, Geophys. J. Roy. Astron. Soc., 16 (1968), pp. 169-205.

Beylkin, G., *The Inversion Problem and Applications of the Generalized Radon Transform*, Comm. on Pure and App. Math., Vol. XXXVII (1984) pp. 579-599.

Beylkin, G., *Imaging of Discontinuities In the Inverse Scattering Problem by Inversion of A Causal Generalized Radon Transform*, J. Math. Phys., 26(1) (1985), pp. 99-108.

Beylkin, G. & Burridge, R., *Linearized Inverse Scattering Problems in Acoustics and Elasticity*, Wave Motion, 12 (1990), pp. 15-52.

Bleistein, N., Cohen, J.K. & Stockwell, J.W.J., *Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion*, Springer-Verlag, New York, 2001.

Borden, B., *Radar Imaging of Airborne Targets: A Premier for Applied Mathematicians and Physicists*, Institute of Physics, Bristol, 1999.

Bydder, M., Samsonov, A.A. & Du, J., *Evaluation of Optimal Density Weighting for Regridding*, Mag. Reson. Imag., 25 (2007), pp. 695-702.

Demmel, J.W., *Applied Numerical Linear Algebra*, Siam, Philadelphia, 1997.

Duistermaat, J.J., *Fourier Integral Operators*, Progress in Mathematics, Birkhäuser, 1995.

Dutt, A. & Rokhlin, V., *Fast Fouier Transforms for Nonequispaced Data*, SIAM J. Sci. Comput., 14 (1993), pp. 1368-1393.

Fang, Y., *Imaging from Sparse Measurements*, Ph.D Thesis, Rensselaer Polytechnic Institute, 2008.

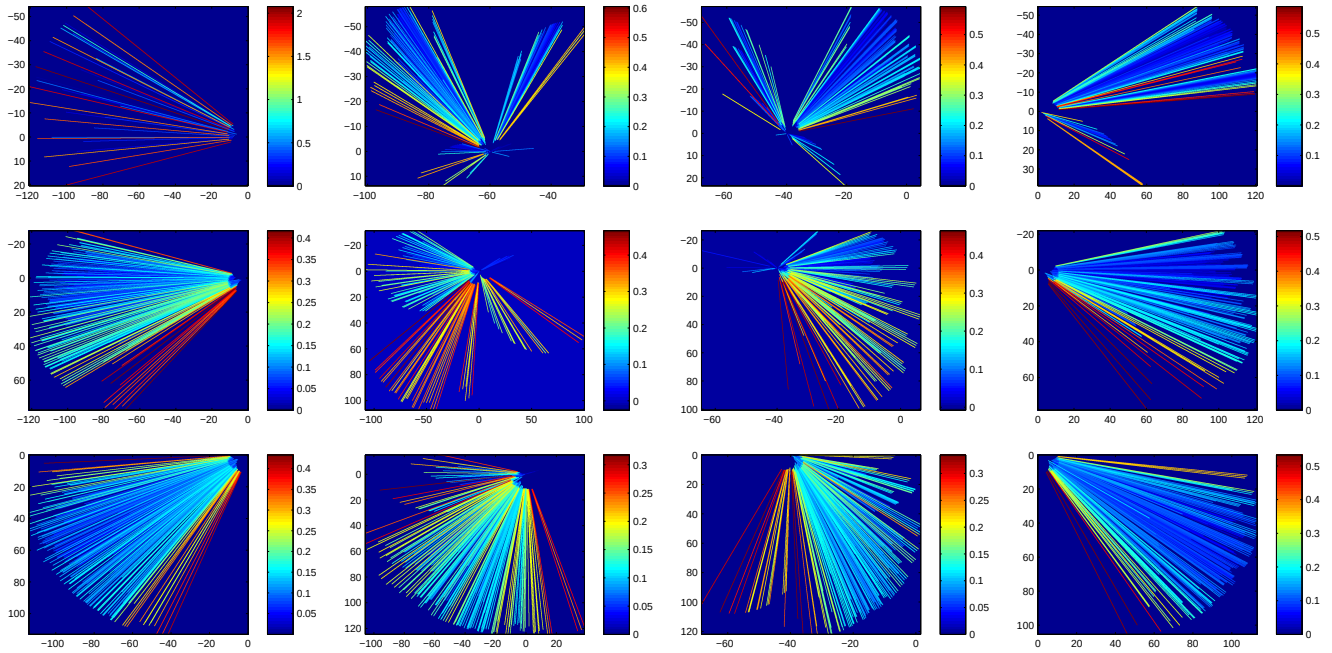


Figure 11. Bisecting line segments plots of 4×3 different imaging points corresponding to the stars in the geometry in Fig. 5. Each bisecting line segment is colored according to the optimization weights. The horizontal axis is k_1 and the vertical axis is k_3. The units on each axis are $1/\text{km}$.

Fessler, J.A. & Sutton, B.P., *Nonuniform Fast Fourier Fransforms Using Min-Max Interpolation*, IEEE Trans. on Sig. Proc, 52 (2003), pp. 560-574.

Friedlander, G. & Joshi, M., *Introduction to the Theory of Distributions, 2nd edition*, Cambridge, Cambridge, 1998.

Greengard, L. & Lee, J., *Accelerating the Nonuniform Fast Fourier Transform*. SIAM Review, 46 (2004), pp 443-454.

Grigis, A. & Sjöstrand, J., *Microloacal Analysis for Differential Operators: An Introduction*, (London Mathematical Society Lecture Note Series vol 196). Cambridge University Press, Cambridge, 1994.

Guillemin, V. & Sternberg, S., *Geometric Asymptotics*, AMS, Providence, 1977.

Jackson, J.D., *Classical Electrodynamics, Third Edition*, Wiley, New York, 1998.

Jackson, J.I., Meyer, D.H. & Nishimura, D.G., *A. Macovski, Selection of A Convolution Function for Fourier Inversion Using Gridding*. IEEE Trans. Med. Imag., 10 (1991), pp. 473-478.

Jakowatz, C.V., Wahl, D.E., Eichel, P.H., Ghiglia, D.C. & Thompson, P.A., *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer, Boston, 1996.

Liang, Z. & Lauterbur, P.C., *Pirnciples of Magmetic Resonance Imaging*, IEEE, New York, 2000.

Malik, W.Q., Khan, H.A., Edwards, D.J. & Stevens, C.J., *A Gridding Algorithm for Efficient Density Compensation of Arbitrarily Sampled Fourier-domain Data*, Advances in Wired and Wireless Communication, (2005), pp. 125-128.

Menke, W., *Geophysical data analysis: discrete inverse theory*, Academic Press, 1989.

Natterer, F. & Wubbeling, F., *Mathematical Methods in Image Reconstruction*, Siam, Philadelphia, 2001.

Nguyen, N. & Liu, Q.H., *The Regular Fourier Matrices and Nonumiform Fast Fourier Transorms*, SIAM J. Sci. Comput., 21 (1999), pp. 283-293.

O'Sullivan, J.D., *A Fast Sinc Function Gridding Algorithm For Fouier Inversion In Computer Tomography*. IEEE Trans. on Med. Imag., MI-4 (1985), pp. 200-207.

Pipe, J.G. & Menon, P., *Smapling Density Compensation in MRI: Rationale and An Iterative Numerical Solution*, Mag. Res. Med., 41 (1999), pp. 179-186.

Pratt, R.G., *Frequency Domain Elastic Wave Modeling by Finite Differences: A Tool for Cross-hole Seismic Imaging*, Geophysics, 55 (1990), pp. 626-632.

Roecker, S., Thurber, C., Roberts, K. & Powell, L., *Refining the Image of the San Andreas Fault Near Parkfield, California Using A Finite Difference Travel Time Computation Technique*, Tectonophysics, 426 (2006), pp. 189-205.

Samsonov, A.A., Kholmovski, E.G. & Johnson, C.R., *Determination of the Sampling Density Compensation Function Using A Point Spread Function Modeling Approach and Gridding Approximation*, Proc. Intl. Soc. Mag. Reson. Med., 11 (2003), pp. 477-477.

Thurber, C., Roecker, S., Roberts, K., Gold, M., Powell, L. & Rittger, K., *Earthquake Locations and Three-dimensional Fault Zone Structure Along the Creeping Section of the San Andreas Fault Near Parkfield, CA, Preparing for SAFOD*, Geophysical Research Letters, 30, 2003.

Tikhonov, A.N. & Arsenin, V.Y., *Solutions of Ill-Posed Problems*, Winston, New York, 1977.

Trefethen, L.N. & Bau, D., III, *Numerical Linear Algebra*, Siam, Philadelphia, 1997.

Wang, Y.H., *Sparseness-constrained least-squares inversion: Application to seismic wave reconstruction*, Geophysics, 2003.

Wang, Y.H. & Rao Y *Reflection Seismic Waveform Tomography*, Journal of Geophysical Research, 114, 2009