Journal of Computational Physics 419 (2020) 109667

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Preconditioner for estimation of multipole sources via full waveform inversion

Mario J. Bencomo *, William W. Symes

The Rice Inversion Project, Rice University, Houston, TX 77005-1892, USA

ARTICLE INFO

Article history:
Received 13 November 2019
Received in revised form 9 June 2020
Accepted 9 June 2020
Available online 15 June 2020

Keywords:
Full waveform inversion
Multipole source
Preconditioning
Conjugate gradient
Ill-conditioned linear inverse problems

ABSTRACT

Accurate representation and estimation of seismic sources is crucial to the joint medium-source full waveform inversion problem. We focus on the source estimation subproblem and its difficulties, where seismic sources are modeled by truncated series of multipoles. The source full waveform inversion formulation results in a highly ill-conditioned (potentially ill-posed) linear least squares problem which we attempt to solve iteratively via conjugate gradient. Our main contribution lies in developing a preconditioner to accelerate the performance of conjugate gradient on multipole source inversion. The proposed preconditioner consists of (fractional) time derivative/integral operators based on analytical solutions to the wave equation with multipole sources in an unbounded, homogeneous medium. Numerical results in 2D demonstrate that the conjugate gradient iterations are accelerated when incorporating the proposed preconditioning scheme.

1. Introduction

Accurate representation and estimation of seismic sources is essential for the recovery of medium parameters in the seismic inverse problem. Joint determination of medium and source parameters via *full waveform inversion* (FWI) has been shown to be successful in many applications as a means of addressing the impact of source parameters on the estimated medium; for instance, Wang et al. [39] and Minkoff and Symes [20]. Actual seismic sources are small relative to other scales in the seismic experiment, and thus are well-modeled as point sources, but also radiate in complex, anisotropic patterns. Anisotropic radiators of point support are necessarily linear combinations of derivatives of the spatial delta function, or *multipole* series, according to Peetre's Theorem [11]; we give a precise definition of multipoles in the theory section. The concept of modeling seismic sources as multipoles is not new. In fact, the multipole representation is a generalization of the *moment tensor* representation of elastic sources commonly used in earthquake seismology (cf. Stump and Johnson [34], Aki and Richards [1], Backus and Mulcahy [2], Shearer [31]).

Regardless of representation, source parameters are seldom measured directly, so must be estimated from data by solution of an inverse problem. We formulate this problem as a linear least squares problem, and observe that the Hessian matrix of the objective functional tends to be ill-conditioned. One reason for this ill-conditioning is the presence of a range of orders of time derivatives of the input source waveforms observed in the data, associated with the spatial derivatives in the multipole series modeling the source. Our main result is to show how this cause of ill-conditioning can be ameliorated by use of appropriately weighted norms. We apply this observation to accelerate the convergence of *conjugate gradient* (CG)

* Corresponding author.
E-mail address: mjb6@rice.edu (M.J. Bencomo).

https://doi.org/10.1016/j.jcp.2020.109667
0021-9991/

iterative solvers for the least squares inverse problem: the norms in which the Hessian becomes better-conditioned generate a preconditioned CG method that is more rapidly convergent than un-preconditioned CG. We illustrate these conclusions with numerical simulations of 2D synthetic seismic propagation from multipole sources, calculated in the time domain using finite difference methods. We are able to recover good fit to data and reasonable multipole series coefficients using CG methods, and in particular we observe a significant acceleration of the reduction of data misfit and inversion errors when preconditioning. While we work entirely in the time domain, equivalent results via CG and preconditioned CG could be obtained in the frequency domain.

We first present a review of prior work on seismic source inversion, as well as joint medium-source estimation, in the context of FWI to motivate our methodology. In the following sections we formulate the multipole source FWI problem and discuss the root causes of its ill-conditioning. Lastly, we motivate and derive our proposed preconditioner and demonstrate its effectiveness in accelerating convergence of CG on some synthetic examples.

2. Literature review

Seismic data contains information about source and receiver responses as well as properties of the earth's medium, the latter being of priority in exploration seismology. Conventional methodology initially focused on removing the source response from data as a preprocessing step in seismic imaging by estimating the source signature through statistical methods like predictive deconvolution [28] and homomorphic deconvolution [37]. Ziolkowski [41] has criticized these statistical methods for imposing unrealistic and at times theoretically unjustifiable constraints on both the source and medium, yielding results vulnerable to subjectivity. Alternatively, source signatures for the *Vibroseis*™ and airgun arrays have been estimated using near-field measurements with some success; see Ziolkowski [41] and Landrø and Sollie [14]. It should be noted that source estimation with near-field measurements still depends on how the source is modeled and the data's dependency on the medium which can be unknown in applications where the source-receiver path is partially submerged in unknown medium or the direct arrival cannot be clearly discerned from the data.

Starting in the late 1980's, efforts in decoupling medium-source interactions shifted from preprocessing the source signature out of seismic traces to the joint inversion of medium and source parameters. Early attempts focused on proving theoretically the co-determinability of source time-dependent parameters and medium parameters from reflectivity data under simplifying assumptions: acoustic layered medium, quasi-impulsive and non-impulsive sources, primaries only data, inverting for only one medium parameter (reflectivity), [26,15,4,19]. Later works have implemented and tested the feasibility of joint medium-source inversion beyond the theory with some success in a variety of synthetic and field examples; [20,39,40]. Some recent works have employed the *variable projection* (VP) method, first developed by Golub and Pereyra [8] (see Golub and Pereyra [9] for a more current discussion on the topic), to exploit the separable structure of joint medium-source inversion. In particular, source parameters are effectively eliminated by an orthogonal projection when solving a linear least squares subproblem. In practice, VP, coupled with Gauss-Newton type algorithms for solving the primary nonlinear least squares problem, has been proven to outperform other multiparameter estimation algorithms, [30]. For example, Rickett [27] and Li et al. [17] demonstrate the effectiveness of VP over simultaneous descent and alternating direction in the context of joint medium-source inversion via FWI in the time and frequency domains respectively.

All of the works mentioned above, with the exception of Minkoff and Symes [20], idealize the source contribution to that of an isotropic source localized at a single point in space, i.e., a point-source. The point-source assumption for seismic sources is justified by the fact that the spatial dimensions of sources considered in exploration seismology are considerably smaller (typically by an order of magnitude) than the propagating wavelengths of seismic waves. The isotropy assumption, however, is questionable, especially when the source is known or potentially expected to exhibit directivity or anisotropy in its radiation pattern. A perfect example is the airgun-array used in marine seismic surveys where the output pressure field of the source is known to exhibit vertical directivity. Work by Minkoff and Symes [20], a prime motivator for this paper, demonstrates the necessity of modeling source anisotropy for airgun-arrays, as well as estimating source parameters in conjunction with the medium for an accurate inversion. Anisotropy of their airgun-array source was modeled in the plane-wave domain by representing the source term as a series of Legendre polynomials in the slowness variable. Their results show that inverting for anisotropic source parameters, as opposed to using a given isotropic source or even inverting for an isotropic source, allowed them to account for 25% more of the data and were even able to achieve a dramatic 90% data fit up to a gas-sand target. Moreover, recovered reflectivity parameters matched closely expected lithology, but only when jointly estimating for medium and anisotropic source parameters.

Active work on anisotropic seismic source inversion primarily has focused on the determination of earthquake mechanisms [32,13], and most recently in the characterization of microseismic events (small earthquakes) resulting from hydraulic fracturing [33,6]. Time-domain methods based on an FWI approach result in having to solve a block Toeplitz system of equations, which can be done efficiently with Levinson recursion algorithms [29,36]. In the frequency domain, the equivalent problem is no longer solved by a simple Fourier division. Instead, one has to solve a system of M equations for M unknowns point-wise in frequency, where M is the number of source components one is inverting for; we discuss this in more detail later. Some of the references cited above consider an alternate frequency-domain method, the so-called *appraisal deconvolution* method by Oldenburg [22]. Appraisal deconvolution consists of estimating, in the frequency domain, a set of optimal inverse filters, rather than source parameters. What sets this method apart from a standard least squares approach in the frequency-domain is the choice of objective function in defining an optimal inverse filter. In the scalar de-

convolution case, one can show equivalency between appraisal deconvolution and a standard least squares approach in time and frequency domain (also equivalent to Wiener deconvolution); this is no longer true for the general multichannel case. Given that an inverse filter is computed, the appraisal deconvolution method readily offers access to qualitative measures of the inversion, primarily the sensitivity matrix. On the other hand, this method requires the careful calibration of several hyper-parameters, and can be computationally intractable for moderate to large scale problems.

A particular challenge of the multipole source inverse problem, as we discuss in the next section, is ill-conditioning. Some attempts at regularization have aimed to improve data coverage, as well as constraining or oversimplifying the source representation. For example, Eaton and Forouhideh [6] and Song and Toksöz [33] invert for time-independent moment tensor components and Eaton and Forouhideh [6] further discusses ill-conditioning of the inverse problem with respect to angle subtended by the receiver array. Koch [13] applies a more standard approach via Tikhonov regularization. Their numerical results indeed demonstrate the benefits of this method, especially in cases where data acquisition geometry results in known ill-conditioning. Furthermore, Koch [13] illustrated robustness of their results against errors in time lag, model parameters (i.e., velocity model), and extended sources (non-point sources). It is worth highlighting that their regularization parameter is chosen in a trial by error manner given information about the singular value decomposition of the forward map, which could be impractical for larger problems. It should also be noted that time-domain algorithms, based on Levinson recursion, are weakly stable [5], making them ill-suited for ill-conditioned problems unless adequate regularization is introduced. The appraisal deconvolution method offers a form of regularization (equivalent to Tikhonov regularization) through a "trade-off" parameter, though there is no efficient manner in setting said parameter.

We pose the multipole source inverse problem in a time-domain FWI framework and solve the resulting linear least squares problem iteratively via CG as done by Wang et al. [39]. An iterative solver based on Krylov subspaces will allow us to efficiently regularize large scale linear inverse problems when using an early termination criterion. Performance issues (i.e., slow convergence) of CG due to ill-conditioning are of particular importance in the multipole source inverse problem and addressed by preconditioning based on analytical solutions to the wave equation with multipole sources. In the following section, we introduce our mathematical models for wave propagation and multipole sources, as well as discuss the analytical motivations for our preconditioning scheme.

3. Theory

We model seismic wave phenomena via the acoustic equations, particularly in the first-order form,

$$\begin{aligned} &\frac{\partial}{\partial t}\mathbf{v}(\mathbf{x},t) + \beta(\mathbf{x})\nabla p(\mathbf{x},t) = 0, \\ &\frac{\partial}{\partial t}p(\mathbf{x},t) + \kappa(\mathbf{x})\nabla\cdot\mathbf{v}(\mathbf{x},t) = f(\mathbf{x},t), \end{aligned} \tag{1}$$

with

- $\beta =$ buoyancy (reciprocal of density),
- $\kappa =$ bulk modulus,
- $p =$ pressure scalar field,
- $\mathbf{v} =$ particle velocity vector field,
- $f =$ (pressure) source term,

for $\mathbf{x} \in \Omega \subset \mathbb{R}^d$ and $t \in [0, T]$, with appropriate boundary and initial conditions and $d \in \{1, 2, 3\}$ denoting the spatial dimension. Note that the source term f is introduced as a scalar in the pressure equation only, though we could have just as easily introduced a vector force in the velocity equations.

In this paper we model general anisotropic sources as *multipole* series involving derivatives of the spatial delta function, e.g.,

$$f(\mathbf{x},t) = \sum_{m=1}^{M} w_m(t) D^{\mathbf{s}_m}\delta(\mathbf{x}-\mathbf{x}^*), \tag{2}$$

for a source centered at $\mathbf{x}^*$. Equation (2) makes use of multi-index notation: given non-negative integer d-tuple $\mathbf{s} = (s_1, \ldots, s_d)$, define the $\mathbf{s}$-mixed partial derivative operator, $D^{\mathbf{s}}$, and its order, $|\mathbf{s}|$, as follows,

$$D^{\mathbf{s}} = \prod_{k=1}^{d}\left(\frac{\partial}{\partial x_k}\right)^{s_k}, \quad |\mathbf{s}| = \sum_{k=0}^{d} s_k.$$

Multipole source f as stated in equation (2), given multi-indexes $\mathbf{s}_m$ and source location $\mathbf{x}^*$, is uniquely determined by its time-dependent functions $w_m(t)$, which we refer to as *multipole coefficients*.

By the superposition principle, we can express the pressure field p resulting from a multipole source f, as given by equation (2), as a temporal multichannel convolution, i.e.,

$$p(\mathbf{x},t) = \sum_{m=1}^{M} g_m(\mathbf{x},t) * w_m(t) \tag{3}$$

where "$*$" denotes convolution in time and g_m is the Green's function associated with acoustic system (1) with source term $\delta(t)D^{\mathbf{s}_m}\delta(\mathbf{x}-\mathbf{x}^*)$ in the pressure equation. We take this multichannel convolution approach when simulating the pressure field during inversion as discussed later.

Consider the following scenario for which analytical solutions can be easily derived. Suppose we have a homogenous and unbounded medium (i.e., β and κ are constants and $\Omega = \mathbb{R}^d$), and that the multipole source is centered at the origin. Let $p_{\mathbf{s}}$ denote pressure field associated with a single-term multipole source given by $f(\mathbf{x},t) = w(t)D^{\mathbf{s}}\delta(\mathbf{x})$, for some given multi-index $\mathbf{s}$ and smooth multipole coefficient $w(t)$. Since we are interested in the pressure field only, we can turn to the well known second-order form of the wave equation; in relation to system (1) and the given single-term multipole source, we have $p_{\mathbf{s}}$ also satisfies

$$\left(\frac{\partial^2}{\partial t^2} - c^2\nabla^2\right)p_{\mathbf{s}}(\mathbf{x},t) = w'(t)D^{\mathbf{s}}\delta(\mathbf{x}),$$

where $c = \sqrt{\kappa\beta}$ is the wave speed. For the case where $\mathbf{s} = \mathbf{0}$, in other words, $D^{\mathbf{s}}\delta(\mathbf{x}) = \delta(\mathbf{x})$, the pressure field is simply given by

$$p_{\mathbf{0}}(\mathbf{x},t) = g(\mathbf{x},t) * w'(t), \tag{4}$$

where $g(\mathbf{x},t)$ is the classic Green's function for the wave equation;

$$g(\mathbf{x},t) = \begin{cases} \dfrac{1}{2c}H(t-\frac{|x|}{c}), & \text{1D}, \\ \dfrac{1}{2\pi x}\dfrac{H(t-\frac{|\mathbf{x}|}{c})}{c^2t^2-|\mathbf{x}|^2}, & \text{2D}, \\ \dfrac{1}{4\pi c^2}\dfrac{\delta(t-\frac{|\mathbf{x}|}{c})}{|\mathbf{x}|}, & \text{3D}. \end{cases} \tag{5}$$

In the above equations, $H(t)$ denotes the Heaviside step function.

Given that spatial derivatives commute with the wave operator (in the case of an unbounded constant medium), it follows that the pressure field for a given multi-index $\mathbf{s}$ is given by

$$p_{\mathbf{s}}(\mathbf{x},t) = D^{\mathbf{s}}p_{\mathbf{0}}(\mathbf{x},t).$$

Table 1 lists examples of $p_{\mathbf{s}}$ for different multi-indexes $\mathbf{s}$ in different dimensions,[1] where

$$\tau = t - \frac{|\mathbf{x}|}{c}, \quad \gamma_k = \frac{x_k}{|\mathbf{x}|}, \quad \xi(\mathbf{x},\sigma) = \left(\sigma^2 + 2\frac{|\mathbf{x}|}{c}\right)^{-1/2},$$
$$\operatorname{sgn}(x) = \begin{cases} -1, & x<0 \\ 0, & x=0 \\ 1, & x>0 \end{cases}, \quad \delta_{kl} = \begin{cases} 1, & k=l \\ 0, & k\neq l \end{cases}.$$

The case where $\mathbf{s} = \mathbf{0}$ results in a pressure field that only depends spatially on $|\mathbf{x}|$. In other words, $f(\mathbf{x},t) = w(t)\delta(\mathbf{x})$ has an isotropic pressure field response; we refer to this source as a *monopole*. Higher order multipoles demonstrate anisotropic pressure field responses evident by terms such as sgn(x) in 1D and γ_k in 2D and 3D. The cases with $|\mathbf{s}| = 1$ and $|\mathbf{s}| = 2$ are referred to as *dipole* and *quadrapole* sources respectively. In 1D, we see that the pressure field is only smooth for the monopole case. In comparison, 2D and 3D analytical solutions are smooth for higher order multipoles as long as $w(t)$ is sufficiently differentiable.

For 1D and 3D it is quite straightforward to make a connection between the multipole coefficient and the pressure field if we focus on the far-field terms, that is, terms that dominate as $|\mathbf{x}| \to \infty$. Effectively, the pressure field consists of a progressive waveform defined by some derivative of the multipole coefficient scaled by powers of $1/c$.[2] The higher the

[1] A change of variables is applied for the 2D case, i.e., $\sigma = \sqrt{t - \frac{|\mathbf{x}|}{c}}$, to obtain a simpler expression. Leibniz rule is applied when differentiating to derive the pressure field with respect to a dipole source.

[2] Note that spatial derivatives on $p_{\mathbf{0}}$ translate to time derivatives and factors of $1/c$ on the multipole coefficient $w(t)$ due to the progressing wave form of the pressure field.

Table 1
Pressure field in 1D, 2D and 3D of the acoustic equations (1) with single-term multipole sources in a homogeneous, unbounded medium.

	Source term f	Analytical solution p
1D	$w(t)\delta(x)$	$\frac{1}{2c}w(\tau)$
	$w(t)\frac{d}{dx}\delta(x)$	$-\frac{1}{2c^2}w'(\tau)\,\mathrm{sgn}(x)$
	$w(t)\frac{d^2}{dx^2}\delta(x)$	$\frac{1}{2c^3}w''(\tau)-\frac{1}{c^2}w'(\tau)\delta(x)$
2D	$w(t)\delta(\mathbf{x})$	$\frac{1}{\pi c^2}\int_0^{\sqrt{\tau}} w'(\tau-\sigma^2)\xi(\mathbf{x},\sigma)\,d\sigma$
	$w(t)\frac{\partial}{\partial x_k}\delta(\mathbf{x})$	$-\frac{\gamma_k}{\pi c^3}\int_0^{\sqrt{\tau}}\left\{w''(\tau-\sigma^2)\xi(\mathbf{x},\sigma)+w'(\tau-\sigma^2)\xi^3(\mathbf{x},\sigma)\right\}d\sigma$
3D	$w(t)\delta(\mathbf{x})$	$\frac{1}{4\pi c^2\lvert\mathbf{x}\rvert}w'(\tau)$
	$w(t)\frac{\partial}{\partial x_k}\delta(\mathbf{x})$	$-\frac{\gamma_k}{4\pi c^2}\left\{\frac{1}{c\lvert\mathbf{x}\rvert}w''(\tau)+\frac{1}{\lvert\mathbf{x}\rvert^2}w'(\tau)\right\}$
	$w(t)\frac{\partial^2}{\partial x_l\partial x_k}\delta(\mathbf{x})$	$\frac{1}{4\pi c^2}\left\{\frac{\gamma_k\gamma_l}{c^2\lvert\mathbf{x}\rvert}w'''(\tau)-\frac{\delta_{kl}-3\gamma_k\gamma_l}{c\lvert\mathbf{x}\rvert^2}w''(\tau)-\frac{\delta_{kl}-3\gamma_k\gamma_l}{\lvert\mathbf{x}\rvert^3}w'(\tau)\right\}$

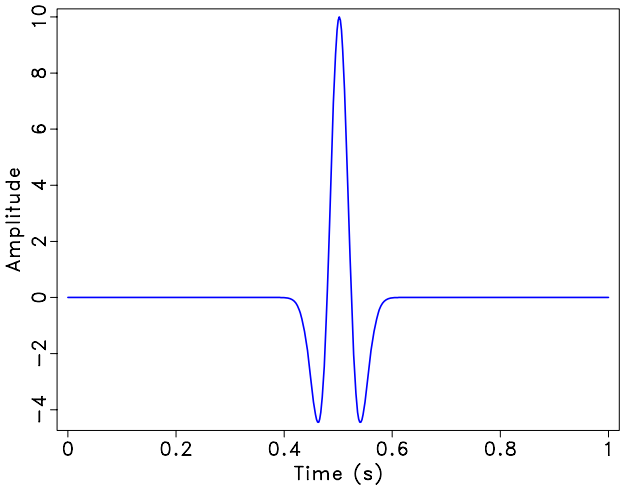


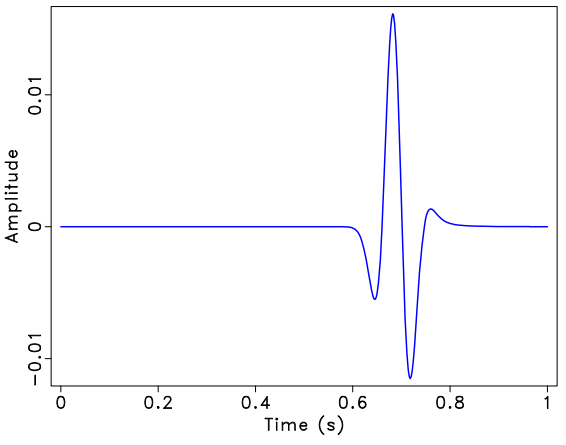
Fig. 1. 10 Hz Ricker wavelet time series.

multipole order, the more derivatives and factors of $1/c$ are tacked onto the corresponding multipole coefficients. In 2D, we still have derivatives of w and scaling by $1/c$, but with the added complexity of an integral. It turns out that the differ-integral form for the 2D analytical solution is quite similar to a (Caputo) fractional derivative; we elaborate more on this when we discuss our preconditioner. As proof of concept, Fig. 2 plots the pressure field as well as fractional time derivatives of the multipole coefficient related to multipole source $f(\mathbf{x},t)=w(t)D^{\mathbf{s}}\delta(\mathbf{x}-\mathbf{x}^*)$ in a 2D "unbounded" constant medium. The multipole coefficient $w(t)$ for these examples was chosen to be a 10 Hz Ricker wavelet, see Fig. 1. Pressure fields and fractional derivates are computed by finite difference methods which we describe in more detail in later sections. Fig. 2 shows almost identical waveforms between pressure fields and fractional derivatives of the multipole coefficient, modulo a scaling factor.

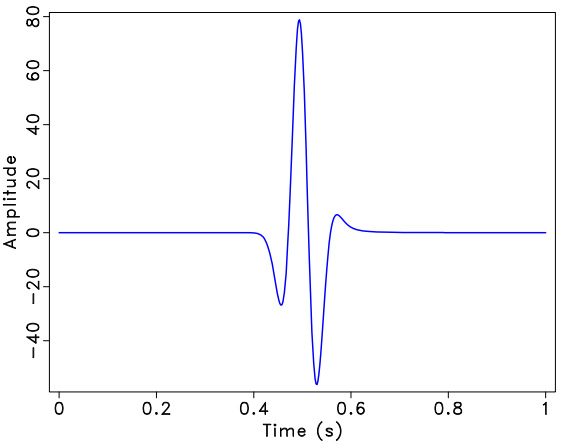
In summary, we observe that the pressure field due to a single-term multipole source, $f(\mathbf{x},t)=w(t)D^{\mathbf{s}}\delta(\mathbf{x}-\mathbf{x}^*)$, is approximately proportional to derivatives of the multipole coefficient. In particular, based on the analytical solutions in Table 1, we have that

$$p(\mathbf{x},t)\approx\frac{1}{c^{|\mathbf{s}|}}\left(\frac{d}{dt}\right)^{\beta+|\mathbf{s}|}w(t-\tfrac{|\mathbf{x}|}{c}) \tag{6}$$

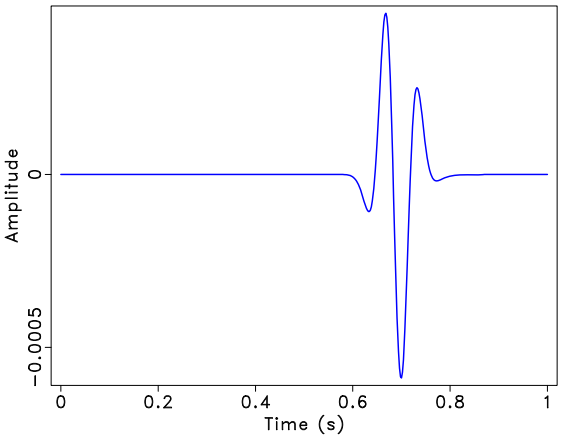
where $\beta=0,\frac{1}{2},1$ for 1D, 2D, and 3D respectively.



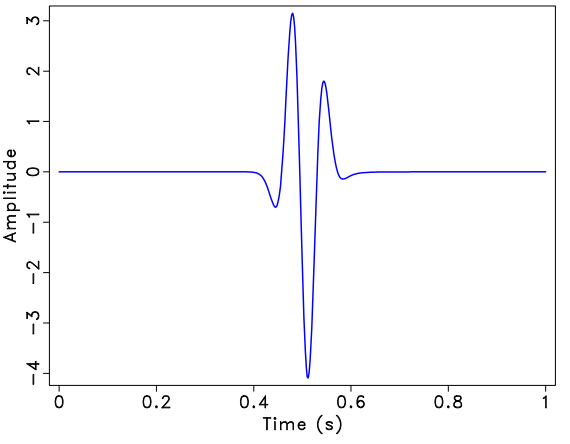
(a) Pressure field for $\mathbf{s} = (0, 0)$.



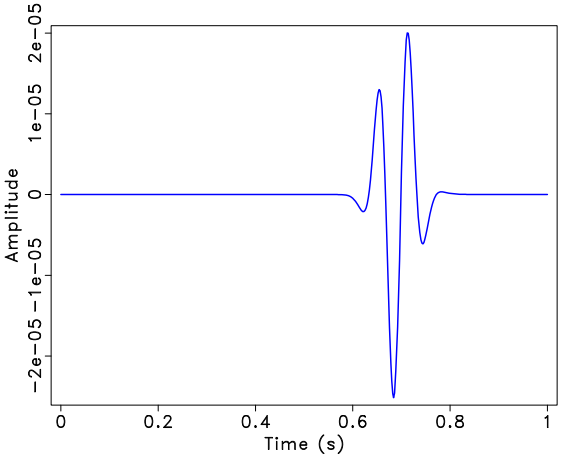
(b) 1/2-derivative of $w(t)$.



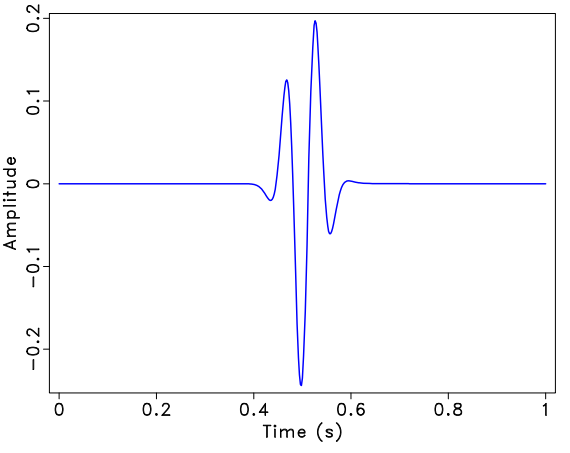
(c) Pressure field for $\mathbf{s} = (0, 1)$.



(d) 3/2-derivative of $w(t)$.



(e) Pressure field for $\mathbf{s} = (0, 2)$.



(f) 5/2-derivative of $w(t)$.

Fig. 2. Comparing pressure field (left column) and fractional derivatives (right column) of the multipole coefficient for a single-term multipole source of the form $f(\mathbf{x}, t) = w(t)D^{\mathbf{s}}\delta(\mathbf{x} - \mathbf{x}^*)$ in 2D. Source is centered at $\mathbf{x}^* = (500\text{ m}, 500\text{ m})$ and pressure field is measured at $\mathbf{x} = (300\text{ m}, 300\text{ m})$.

4. FWI formulation

We turn to formulating the inverse problem in this section. Again, our goal is to determine the multipole source that best fits some "observed" pressure field data. The source is modeled as a multipole series given by equation (2) and the data is generated by solving the acoustic equations (1) at specified receiver locations, with said source. We assume that the source representation is given, that is, we know the source location $\mathbf{x}^*$ and the multi-indexes $\mathbf{s}_m$. Thus, the multipole source inverse problem consists of determining multipole coefficients $w_m(t)$.

We introduce the following notation in formulating the inverse problem. Multipole coefficients are grouped to form the *multipole coefficient vector* $\mathbf{w}$,

$$\mathbf{w} = \{w_m(t) : m = 1, ..., M\}, \tag{7}$$

which has a natural vector space structure if we endow such vectors with the following obvious definition of linear combinations: for two multipole coefficient vectors $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$, and scalars a_1, a_2,

$$a_1\mathbf{w}^{(1)} + a_2\mathbf{w}^{(2)} = \{a_1 w_m^{(1)}(t) + a_2 w_m^{(2)}(t) : m = 1, ..., M\}.$$

The space of multipole coefficient vectors, $\mathbb{W}$, is referred to as the *multipole source (MPS) space*. Note that we regard source position $\mathbf{x}^*$ and multi-indexes $\mathbf{s}_m$ as attributes of an MPS space. We also define a natural L^2 inner-product for MPS spaces as follows,

$$\langle \mathbf{w}^{(1)}, \mathbf{w}^{(2)} \rangle_{\mathbb{W}} := \sum_{m=1}^{M} \int w_m^{(1)}(t) w_m^{(2)}(t)\, dt. \tag{8}$$

We denote $\|\cdot\|_{\mathbb{W}}$ as the norm induced by the inner product above, and refer to it as the L^2 norm for $\mathbb{W}$.

The data we consider in our inversions consists of time traces of the pressure field measured at discrete receiver locations. Similarly to the multipole coefficient vector, we group pressure field traces sampled at receiver points $\{\mathbf{x}_n : n = 1, 2, ..., N\}$ to form the *data vector* $\mathbf{d}$,

$$\mathbf{d} = \{p(\mathbf{x}_n, t) : n = 1, ..., N\}.$$

We let $\mathbb{D}$ denote the space of data vectors, or simply the *data space*, which we also equip with a standard L^2 inner-product,

$$\langle \mathbf{d}^{(1)}, \mathbf{d}^{(2)} \rangle_{\mathbb{D}} := \sum_{n=1}^{N} \int p^{(1)}(\mathbf{x}_n, t) p^{(2)}(\mathbf{x}_n, t)\, dt. \tag{9}$$

Denote $\|\cdot\|_{\mathbb{D}}$ as the norm induced by the inner product above, and refer to it as the L^2 norm for $\mathbb{D}$.

Define the *forward map* as an operator $\mathbf{F} : \mathbb{W} \to \mathbb{D}$ from MPS space $\mathbb{W}$ to data space $\mathbb{D}$, e.g., $\mathbf{Fw} = \mathbf{d}$ with $\mathbf{w} \in \mathbb{W}$ and $\mathbf{d} \in \mathbb{D}$. As discussed in the theory section, the forward map is a linear mapping consisting of a multichannel convolution between the multipole coefficients and an appropriate set of Green's functions. In particular, we can view $\mathbf{F}$ as an $N \times M$ matrix of linear operators,

$$\mathbf{F} = \begin{bmatrix} F_{11} & \cdots & F_{1M} \\ \vdots & \ddots & \vdots \\ F_{N1} & \cdots & F_{NM} \end{bmatrix}, \quad (F_{nm} w_m)(t) = g_m(\mathbf{x}_n, t) * w_m(t)$$

where g_m are the same Green's functions used in equation (3). Hence, given $\mathbf{w} \in \mathbb{W}$, we have that

$$\mathbf{Fw} = \left\{ \sum_{m=1}^{M} (F_{nm} w_m)(t) \ : \ n = 1, ..., N \right\} = \left\{ \sum_{m=1}^{M} g_m(\mathbf{x}_n, t) * w_m(t) \ : \ n = 1, ..., N \right\}. \tag{10}$$

In practice the number of receivers N will exceed the number of multipole terms M. For numerical results presented here, $N = 200$ while $M = 3$ at most, which implies that $\mathbf{Fw} = \mathbf{d}_{\text{obs}}$, where $\mathbf{d}_{\text{obs}}$ denotes the observed data, is an overdetermined system. Moreover, observed data will be polluted by noise and thus we cannot expect nor want to find an exact solution to this linear problem. Thus we turn to a least squares formulation of the inverse problem. The multipole source FWI problem consists solving the following minimization problem:

$$\min_{\mathbf{w} \in \mathbb{W}} \frac{1}{2} \|\mathbf{Fw} - \mathbf{d}_{\text{obs}}\|_{\mathbb{D}}^2. \tag{11}$$

Solution to minimization problem (11) is given by the solution to the normal equations:

$$\mathbf{F}^T\mathbf{F}\mathbf{w} = \mathbf{F}^T\mathbf{d}_{\text{obs}}, \tag{12}$$

with $\mathbf{F}^T$ denoting the adjoint of $\mathbf{F}$ with respect to source parameter and data spaces and their respective inner products. Given the multichannel convolution representation of the forward map, equation (10), and the standard choice of L^2 inner-products, it follows that $\mathbf{F}^T$ takes the form of a multichannel cross-correlation. In particular, let $\mathbf{d} = \{p(x_n, t) \,:\, n = 1, ..., N\}$ be some vector in the data space, then

$$\mathbf{F}^T\mathbf{d} = \left\{ \sum_{n=1}^{N} g_m(\mathbf{x}_n, t) \star p(x_n, t) \,:\, m = 1, ..., M \right\}$$

where "$\star$" denotes cross-correlation in time.

To paraphrase our discussion of analytical solutions to the wave equation with multipole sources in a constant, unbounded medium, the pressure field due to a multipole source resembles that of time derivatives of the multipole coefficients; see equation (6). In other words, $\mathbf{F}$ resembles a differential operator. Moreover, given the standard choice of L^2 inner product for the domain of the forward map, we obtain an unbounded linear operator which in turn results in an ill-conditioned inverse problem in both the continuum and discrete settings. We explore this source of ill-conditioning in more detail with the following example.

Consider least squares problem (11) in 1D over a constant and unbounded medium with multipole source

$$f(x, t) = w_1(t)\delta(x) + w_2(t)\frac{d\delta}{dx}(x). \tag{13}$$

To simplify the analysis we pose the inversion in the frequency domain. Applying Plancherel's theorem results in the following equivalent least squares problem:

$$\min_{\widehat{\mathbf{w}}} \frac{1}{2}\|\widehat{\mathbf{F}}\widehat{\mathbf{w}} - \widehat{\mathbf{d}}_{\text{obs}}\|^2_{\mathbb{D}} \tag{14}$$

where "$\widehat{\ }$" denotes the Fourier transform of a given function. For example,

$$\widehat{w}_m(\omega) = \int_{\mathbb{R}} w_m(t) e^{-2\pi i t\omega}\, dt.$$

It follows from the convolution theorem that the forward map in the frequency domain, $\widehat{\mathbf{F}}$, consists of simply matrix multiplication, point-wise in frequency. In particular,

$$\widehat{\mathbf{F}}(\omega) = \begin{bmatrix} \widehat{g}_1(x_1, \omega) & \cdots & \widehat{g}_M(x_1, \omega) \\ \vdots & \ddots & \vdots \\ \widehat{g}_1(x_N, \omega) & \cdots & \widehat{g}_M(x_N, \omega) \end{bmatrix} \in \mathbb{C}^{N\times M},$$

and thus

$$(\widehat{\mathbf{F}}\widehat{\mathbf{w}})(\omega) = \left\{ \sum_{m=1}^{M} \widehat{g}_m(x_n, \omega)\widehat{w}_m(\omega) \,:\, n = 1, ..., N \right\}.$$

Solution to minimization problem (14) is given by the solution to the following normal equation, analogous to that of (12), though point-wise in frequency:

$$(\widehat{\mathbf{F}}^H\widehat{\mathbf{F}}\widehat{\mathbf{w}})(\omega) = (\widehat{\mathbf{F}}^H\widehat{\mathbf{d}}_{\text{obs}})(\omega)$$

where $\widehat{\mathbf{F}}^H$ denotes the conjugate transpose of $\widehat{\mathbf{F}}$.

For multipole source (13), we have $M = 2$ and

$$g_1(x, t) = \frac{1}{2c}\delta(t - \tfrac{|x|}{c}) \quad \Longrightarrow \quad \widehat{g}_1(x, \omega) = \frac{1}{2c}\exp(-\tfrac{2\pi i\omega}{c}|x|)$$

$$g_2(x, t) = \frac{1}{2c^2}\delta'(t - \tfrac{|x|}{c})\text{sgn}(x) \quad \Longrightarrow \quad \widehat{g}_2(x, \omega) = \frac{\pi i\omega}{c^2}\exp(-\tfrac{2\pi i\omega}{c}|x|)\,\text{sgn}(x)$$

in accordance to Table 1. Thus,

$$\widehat{\mathbf{F}}(\omega) = \begin{bmatrix} \mathbf{f}_1(\omega), & \mathbf{f}_2(\omega) \end{bmatrix} \in \mathbb{C}^{N\times 2}$$

with column vectors

$$\mathbf{f}_1(\omega) = \frac{1}{2c}\begin{bmatrix} \exp(-\frac{2\pi i\omega}{c}|x_1|) \\ \vdots \\ \exp(-\frac{2\pi i\omega}{c}|x_N|) \end{bmatrix}, \quad \mathbf{f}_2(\omega) = \frac{\pi i\omega}{c^2}\begin{bmatrix} \exp(-\frac{2\pi i\omega}{c}|x_1|)\,\mathrm{sgn}(x_1) \\ \vdots \\ \exp(-\frac{2\pi i\omega}{c}|x_N|)\,\mathrm{sgn}(x_N) \end{bmatrix}.$$

Moreover, the matrix $\widehat{\mathbf{F}}^H\widehat{\mathbf{F}}$ is given by

$$\begin{aligned}(\widehat{\mathbf{F}}^H\widehat{\mathbf{F}})(\omega) &= \begin{bmatrix} (\mathbf{f}_1^H\mathbf{f}_1)(\omega) & (\mathbf{f}_1^H\mathbf{f}_2)(\omega) \\ (\mathbf{f}_2^H\mathbf{f}_1)(\omega) & (\mathbf{f}_2^H\mathbf{f}_2)(\omega) \end{bmatrix} \\ &= \frac{1}{4c^2}\begin{bmatrix} N & \frac{2\pi i\omega}{c}\sum_{i=1}^N \mathrm{sgn}(x_i) \\ -\frac{2\pi i\omega}{c}\sum_{i=1}^N \mathrm{sgn}(x_i) & \frac{4\pi^2\omega^2}{c^2}N \end{bmatrix} \in \mathbb{C}^{2\times 2}.\end{aligned}$$

Consider the case where all of the receiver points $\{x_i : i = 1, ..., N\}$ lie either to the right or left of the source, resulting in

$$\sum_{i=1}^N \mathrm{sgn}(x_i) = N \quad \text{or} \quad -N$$

respectively. For this case, we have that $\widehat{\mathbf{F}}^H\widehat{\mathbf{F}}(\omega)$ is rank deficient. In other words, such data coverage fails to capture any of the anisotropy, making the monopole term (isotropic source term) indistinguishable from the dipole term. This placement of receiver points constitutes the worst-case scenario, resulting in an ill-posed inverse problem. The optimal data-coverage scenario consists of having half of the receiver points on the left and right of the source. Assuming we have an even number of receiver points, this leads to $\sum_{i=1}^N \mathrm{sgn}(x_i) = 0$, and

$$(\widehat{\mathbf{F}}^H\widehat{\mathbf{F}})(\omega) = \frac{1}{4c^2}\begin{bmatrix} N & 0 \\ 0 & \frac{4\pi^2\omega^2}{c^2}N \end{bmatrix}.$$

Matrix $\widehat{\mathbf{F}}^H\widehat{\mathbf{F}}$ is clearly invertible in the case of optimal data-coverage, though with a condition number that is quadratic in the wavelength, i.e.,

$$\kappa((\widehat{\mathbf{F}}^H\widehat{\mathbf{F}})(\omega)) = \mathcal{O}(\tfrac{\omega^2}{c^2}) = \mathcal{O}(\lambda^{-2}). \tag{15}$$

Given our discussion of analytical solutions of the wave equation with multipole sources, we can extrapolate result (15) to cases with higher order multipole sources. Recall that every extra spatial derivative on the delta function in the multipole source incurs an added time derivative on the multipole coefficient observed in the pressure field, which translates to an extra factor of ω/c in the frequency domain. Suppose source term f contains a monopole term and that the highest order multipole is of order M^*, then we have that

$$\kappa((\widehat{\mathbf{F}}^H\widehat{\mathbf{F}})(\omega)) = \mathcal{O}(\lambda^{-2M^*}).$$

Moreover, given that the Fourier transform is a unitary operator, we can conclude that in time domain, $\mathbf{F}^T\mathbf{F}$ has condition number $\mathcal{O}(\lambda_{\min}^{-2M^*})$, where $\lambda_{\min}$ is the smallest wavelength. Note that in the discrete setting, where time dependent quantities are time sampled and thus $\lambda_{\min} > 0$, our analysis above implies that the condition number of the forward map increases as the time grid is refined, which is characteristic of unbounded operators.

Clearly, data coverage has a direct impact on the well-posedness and conditioning of the inverse problem. However, the analysis above reveals that multipole source inversion is inherently ill-conditioned, even under ideal source-receiver configurations. This ill-conditioning stems from the singular nature of the multipole source terms, which translates into a pressure field response as derivatives of the input waveform $\mathbf{w}$.

5. Preconditioner

The approach developed here seeks to better condition multipole source inversion by redefining the MPS space $\mathbb{W}$ to yield a better bounded operator $\mathbf{F}$, hence improving the condition number associated with solving the normal equations. We redefine $\mathbb{W}$ by replacing its native inner-product $\langle\cdot,\cdot\rangle_{\mathbb{W}}$ by a weighted one denoted by $\langle\cdot,\cdot\rangle_{\mathbb{W}'}$,

$$\langle\mathbf{w}^{(1)}, \mathbf{w}^{(2)}\rangle_{\mathbb{W}'} := \langle\mathbf{L}\mathbf{w}^{(1)}, \mathbf{L}\mathbf{w}^{(2)}\rangle_{\mathbb{W}} = \langle\mathbf{w}^{(1)}, \mathbf{L}^T\mathbf{L}\mathbf{w}^{(2)}\rangle_{\mathbb{W}}.$$

The norm induced by this weighted inner-product is denoted by $\|\cdot\|_{\mathbb{W}'}$ and simply referred to as the weighted norm. Note that we can express the adjoint of $\mathbf{F}$ with respect to our weighted inner-product $\langle\cdot,\cdot\rangle_{\mathbb{W}'}$ in terms of $\mathbf{F}^T$, namely

$$\mathbf{F}^T \to (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{F}^T.$$

The least squares problem with respect to our new inner-product results in the following modified normal equations,

$$\mathbf{M}^{-1}\mathbf{F}^T\mathbf{F}\mathbf{w} = \mathbf{M}^{-1}\mathbf{F}^T\mathbf{d}_{\text{obs}}, \tag{16}$$

where we have effectively (left) preconditioned by $\mathbf{M} = \mathbf{L}^T\mathbf{L}$. Implementation wise, we take the preconditioning approach instead of actually weighting the MPS space inner product.

Effectiveness of our preconditioner hinges on a proper choice of $\mathbf{L}$: a linear operator that is numerically stable, has an inverse we can readily implement, and most importantly better-conditions our normal equations. Based on our discussion of analytical solutions to the wave equation with multipole sources, in particular equation (6), we choose $\mathbf{L}$ to be a diagonal matrix of linear operators, i.e.,

$$\begin{aligned}
&\mathbf{L}\mathbf{w} = \{(L_m w_m)(t) \,:\, m = 1, ..., M\} \\
&L_m = \frac{1}{c^{|\mathbf{s}_m|}}\left(\frac{d}{dt}\right)^{\beta+|\mathbf{s}_m|}, \quad \beta = \begin{cases} 0, & \text{in 1D}, \\ \frac{1}{2}, & \text{in 2D}, \\ 1, & \text{in 3D}, \end{cases}
\end{aligned} \tag{17}$$

where $c = \sqrt{\kappa\beta}$ is the ambient wave speed.

Note that in 2D, L_m corresponds to a fractional derivative. We follow the Grünwald-Letnikov definition of the left fractional derivative since it naturally leads to a finite-difference-like discretization; see Li and Zeng [16] for a detailed account of numerical methods for fractional calculus. The (left) fractional derivative of order $\alpha > 0$ for a given function $u(t)$, $t \in (0, T]$, is defined as

$$\left(\frac{d}{dt}\right)^{\alpha} u(t) := \lim_{\substack{\Delta t \to 0 \\ N\Delta t = t}} \frac{1}{\Delta t^{\alpha}} \sum_{n=0}^{N} (-1)^n \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-n+1)n!} u(t - n\Delta t). \tag{18}$$

We discretize the fractional derivative by simply not taking the limit in equation (18). Suppose we discretize the time interval $[0, T]$ by introducing $K+1$ uniformly spaced time grid points t_k, with

$$t_k = k\Delta t, \quad k = 0, 1, ..., K,$$

where $\Delta t = \frac{T}{K}$, and let $\mathbf{u} = [u(x_0), ..., u(x_K)]^T$. The discrete fractional derivative of order $\alpha > 0$, is denoted by the matrix $\mathbf{D}_t^{\alpha}$, with

$$(\mathbf{D}_t^{\alpha}\mathbf{u})_k = \frac{1}{\Delta t^{\alpha}} \sum_{n=0}^{N^*} (-1)^n \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-n+1)n!} u(t_{k-n}), \quad N^* = \begin{cases} \min(k, \alpha), & \alpha \in \mathbb{N}, \\ k, & \text{otherwise}. \end{cases} \tag{19}$$

The inverse of a fractional derivative coincides with a fractional integral. The discrete fractional integral of order $\alpha > 0$ is denoted by the matrix $\mathbf{D}_t^{-\alpha}$, with

$$(\mathbf{D}_t^{-\alpha}\mathbf{u})_k = \Delta t^{\alpha} \sum_{n=0}^{k} \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)n!} u(t_{k-n}). \tag{20}$$

It can be shown that whenever $\alpha \in \mathbb{N}$, then

$$\mathbf{D}_t^{\alpha} = (\mathbf{D}_t^1)^{\alpha}, \quad \mathbf{D}_t^{-\alpha} = (\mathbf{D}_t^{-1})^{\alpha}$$

where $\mathbf{D}_t^1$ coincides with a backward finite difference approximation to the first derivative and $\mathbf{D}_t^{-1}$ is related to the left Riemann sum approximation to integration. Moreover, for $\alpha > 0$,

$$\mathbf{D}_t^{-\alpha} = (\mathbf{D}_t^{\alpha})^{-1}.$$

Thus, in computing the action of $\mathbf{L}^{-1}$ or $\mathbf{L}^{-T}$, we work with fractional integrals as given by the discrete operator $\mathbf{D}_t^{-\alpha}$ for some $\alpha > 0$, and its transpose.

6. Numerical details

Up to now, our presentation of the multipole source inverse problem has been stated in the continuum, dealing with functions of time. We fully discretize said inverse problem by simply introducing uniform time grids. We re-interpret our established notation to denote fully discretized quantities as follows. Consider a general MPS space $\mathbb{W}$, which essentially consists of M-dimensional vector-valued time functions, as defined by equation (7). Alternatively, we reuse $\mathbb{W}$ to denote $\mathbb{R}^{MK_w}$, where we assume that each associated multipole coefficient has been discretized over the same temporal grid with

K_w points, and we have concatenated each of the discretized multipole coefficients into a single column vector of size MK_w. Similarly, the discrete analog of a vector in $\mathbb{D}$ is a concatenated vector in $\mathbb{R}^{NK_d}$. Note that we have introduced two number of time points, K_w and K_d for multipole coefficients and data vectors respectively. This is meant to imply the possibility of different time grids, though we will assume that the same grid size Δt is used. In practice $K_d > K_w$ since the data must record the entirety of the source response as well as take into account travel-time between source and receivers. Lastly, the discrete L^2 inner-products are given by

$$\langle \mathbf{w}^{(1)}, \mathbf{w}^{(2)} \rangle_{\mathbb{W}} = \Delta t \sum_{m=1}^{M} \sum_{k=1}^{K_w} w_m^{(1)}(t_k)\, w_m^{(2)}(t_k)$$

$$\langle \mathbf{d}^{(1)}, \mathbf{d}^{(2)} \rangle_{\mathbb{D}} = \Delta t \sum_{n=1}^{N} \sum_{k=1}^{K_d} p^{(1)}(x_n, t_k)\, p^{(2)}(x_n, t_k).$$

In the discrete setting, the forward map $\mathbf{F}$ corresponds to a block Toeplitz matrix of size $(NK_d) \times (MK_w)$, though it is never constructed explicitly. When generating observed data, the pressure field is computed directly via a PDE solver. During inversion, however, the forward map is implemented as a multichannel convolution as suggested by equation (10). Kernels required in the discrete multichannel convolution, i.e.,

$$g_m(x_n, t_k), \quad n = 1, ..., N, \quad m = 1, ..., M, \quad k = 1, ..., K_d, \tag{21}$$

are computed once at the beginning of the inversion by solving acoustic system (1) with appropriate impulsive multipole sources, for a total of M-PDE solves. The memory cost associated in storing these kernels is that of M data vectors, as suggested by equation (21).

The underlying PDE solver used here is a staggered-grid finite difference method of second-order in time and fourth-order in space, also referred to as the 2-4 scheme, based on Madariaga [18] for elasticity. It is worth mentioning that special care is taken when computing Green's functions via finite difference solvers since it involves solving a wave equation with highly singular source terms. We refer to Bencomo and Symes [3] for details pertaining to singular source discretizations for first order wave equations such as acoustic system (1) via staggered-grid finite difference schemes. This approach to singular source discretization is based on work by Waldén [38], where singularities are replaced by regular functions of compact support that satisfy vanishing discrete moment conditions. Such discretizations of singular source terms have been shown to be accurate by preserving convergence rates of finite difference solvers; for example see Waldén [38] and Petersson and Sjögreen [25] for numerical results related to 1D Helmholtz and 2D elastic wave equation respectively. In particular, since we are using a fourth-order finite difference method in space we also use a fourth-order approximation for singular source terms.

As discussed in previous sections, the multipole source inverse problem, when formulated as a linear least squares problem, results in having to solve normal equations (12). We solve these normal equations iteratively via conjugate gradient for normal equations (CGLS), both with and without preconditioner $\mathbf{M} = \mathbf{L}^T\mathbf{L}$ consisting of (potentially fractional) derivative operators. We refer to Paige and Saunders [24] for a description of the CGLS algorithm and Nocedal and Wright [21] for the preconditioned variant which we refer to as PCGLS.

A robust set of stopping criteria is important for any CG algorithm, but especially in the context of ill-conditioned, linear inverse problems; early termination can provide regularization effects but at the price of a higher data misfit. Given the intended scope of this paper, however, we use the following, albeit crude, stopping criterion in order to highlight the acceleration achieved by our preconditioners.

Stopping criterion: Stop if either of the following three conditions are satisfied:

- reached maximum number of iterations, $k = 150$
- sufficient reduction in residual, $\frac{\|\mathbf{r}^{(k)}\|_{\mathbb{D}}}{\|\mathbf{r}^{(0)}\|_{\mathbb{D}}} \leq 10^{-3}$
- sufficient reduction in normal residual, $\frac{\|\mathbf{g}^{(k)}\|_{\mathbb{W}}}{\|\mathbf{g}^{(0)}\|_{\mathbb{W}}} \leq 10^{-5}$

where $\mathbf{r}^{(k)} = \mathbf{d}_{\text{obs}} - \mathbf{F}\mathbf{w}^{(k)}$ is the residual (or data misfit) at the kth CG iterate $\mathbf{w}^{(k)}$. The normal residual, $\mathbf{g}^{(k)}$, is defined as the residual to the normal equations which differs for different formulations:

$$\mathbf{g}^{(k)} = \begin{cases} \mathbf{F}^T\mathbf{r}^{(k)}, & \text{CGLS}, \\ \mathbf{M}^{-1}\mathbf{F}^T\mathbf{r}^{(k)}, & \text{PCGLS} \end{cases}$$

We note that for all of the numerical results shown in the following section, we start our CG algorithms with a zero initial estimate, $\mathbf{w}^{(0)} = 0$, hence $\mathbf{r}^{(0)} = \mathbf{d}_{\text{obs}}$.

7. Numerical experiments

In this section we present 2D source inversion results illustrating the better-conditioning of normal equations (12) and (24) when using our preconditioning scheme. The finite difference solver, multipole source representation and discretization, and inversion framework for the multipole source inverse problem have been implemented as part of the *IWave* and the *Rice Vector Library* (RVL) packages; see Symes et al. [35] and Padula et al. [23] for more information on the related software.

Spatial grid size, Δx_{FD}, for the finite difference solver is chosen so that there are at least 5 grid points per wavelength. Specifically, we take

$$\Delta x_{\text{FD}} \leq \frac{\lambda_{\min}}{5} = \frac{c_{\min}/\omega_{\max}}{5}$$

where $c_{\min}$ is the smallest wave speed in the medium and $\omega_{\max}$ is the maximum frequency of the source. For numerical experiments presented here, $c_{\min} = 1.5$ km/s and $\omega_{\max} \approx 30$ Hz, thus $\Delta x_{\text{FD}} \leq 10$ m. In general,[3]

$$\Delta x_{\text{FD}} = \begin{cases} 5 \text{ m}, & \text{when computing observed data,} \\ 10 \text{ m}, & \text{when computing Green's functions during inversion.} \end{cases}$$

Gaussian noise is also added to observed data as follows. Let $\mathbf{d}^*_{\text{obs}}$ denote the noise-free observed data computed. Gaussian noise ϵ is added,

$$\mathbf{d}_{\text{obs}} = \mathbf{d}^*_{\text{obs}} + \epsilon$$

such that

$$\frac{\text{RMS}(\epsilon)}{\text{RMS}(\mathbf{d}^*_{\text{obs}})} = \frac{\|\epsilon\|_{\mathbb{D}}}{\|\mathbf{d}^*_{\text{obs}}\|_{\mathbb{D}}} = \eta,$$

for a prescribed noise level η.

We note that the time step size used by the finite difference solver may differ from the Δt used to represent discretized quantities in $\mathbb{W}$ and $\mathbb{D}$. In particular, we use a uniform sampling rate of $\Delta t = 0.002$ s in all of the test cases and for all time-dependent quantities, i.e., multipole coefficients, data vectors, and computed Green's functions. The time step size for the finite difference solver, which we denote by Δt_{FD}, may be different from Δt since it must satisfy the following CFL condition,

$$\Delta t_{\text{FD}} < \frac{0.606}{c_{\max}} \Delta x.$$

For the first three tests, where $c_{\max} = 1.5$ km/s, we take

$$\Delta t_{\text{FD}} = \begin{cases} 0.002 \text{ s}, & \Delta x = 10 \text{ m}, \\ 0.001 \text{ s}, & \Delta x = 5 \text{ m}. \end{cases}$$

In tests 4 and 5, where $c_{\max} = 4.5$ km/s, we use

$$\Delta t_{\text{FD}} = \begin{cases} 0.001 \text{ s}, & \Delta x = 10 \text{ m}, \\ 0.0005 \text{ s}, & \Delta x = 5 \text{ m}. \end{cases}$$

As previously mentioned, we solve normal equations (12) via CGLS, with and without preconditioning. Our stopping criteria is based solely on sufficient reduction in residuals or normal residuals, or a maximum number of iterations. Given that we have access to true solution, denoted $\mathbf{w}_{\text{true}}$, we will also keep track of the normalized inversion error,

$$\frac{\|\mathbf{w}^{(k)} - \mathbf{w}_{\text{true}}\|}{\|\mathbf{w}_{\text{true}}\|}$$

with respect to the standard L^2 norm $\|\cdot\|_{\mathbb{W}}$, as well as the weighted norm $\|\cdot\|_{\mathbb{W}'}$ for PCGLS. Ideally, our source inversion will result in estimated multipole coefficients that reduce the residual to noise level and an estimate that is close to the true solution.

The first three tests we present involve estimating different multipole sources in a 2D constant medium with "unbounded" domain simulated using *perfectly matched layers* (PMLs); see Hu et al. [12] for details on the PML implementation taken here. The physical domain is given by $\Omega = [0, 1] \times [0, 1]$ in units of kilometers, with $\kappa = 2.25$ GPa and $\beta = 1\text{cm}^3/\text{g}$" and thus

[3] A smaller Δx_{FD} is taken in order to avoid any potential inverse crimes.

a wave speed of $c = 1.5$ km/s. A single multipole source is positioned at the center of the domain, that is, $\mathbf{x}^* = (500, 500)$ where components x_1 and x_2 refer to depth and lateral/horizontal axes respectively in meters. Two line of receivers, one horizontal and one vertical, are used in order to capture the full anisotropy of the radiating source. Using Matlab notation, the two lines of receivers are given by $(10, 10 : 10 : 990)$ and $(10 : 10 : 990, 10)$.

Test 1: Estimating monopole source

In test 1 we estimate multipole coefficient vector $\mathbf{w} = w_1$ for monopole source

$$f(\mathbf{x}, t) = w_1(t)\delta(\mathbf{x} - \mathbf{x}^*).$$

According to equation (17) the corresponding (discrete) weight operator $\mathbf{L}$, associated with our preconditioner $\mathbf{M} = \mathbf{L}^T\mathbf{L}$, is given by the fractional 1/2-derivative,

$$\mathbf{L} = \mathbf{D}_t^{1/2},$$

computed according to equation (19). The inverse of this operator, $\mathbf{L}^{-1}$, as required for PCGLS, is given by the fractional 1/2-integral,

$$\mathbf{L}^{-1} = \mathbf{D}_t^{-1/2},$$

computed according to equation (20). Fig. 5a plots the noise-free data $\mathbf{d}^*_{\text{obs}}$; a noise level of $\eta = 20\%$ is added to generate observed data $\mathbf{d}_{\text{obs}}$. Again, the observed data consists of time traces of the pressure field observed along two lines of receivers. The first 100 receivers correspond to the line of receivers $(10, 10 : 10 : 990)$, parallel to the horizontal/x_2 axis. The second set of 100 receivers make up the other line $(10 : 10 : 990, 10)$, parallel to the depth/x_1 axis. As expected a single waveform, the same for across both line of receivers, is observed.

Fig. 6 plots the log of the relative norm of the residual (or data misfit) $\mathbf{r}^{(k)}$ versus CG iteration index k, i.e.,

$$\log \frac{\|\mathbf{r}^{(k)}\|_{\mathbb{D}}}{\|\mathbf{r}^{(0)}\|_{\mathbb{D}}} = \log \frac{\|\mathbf{d}_{\text{obs}} - \mathbf{F}\mathbf{w}^{(k)}\|_{\mathbb{D}}}{\|\mathbf{d}_{\text{obs}}\|_{\mathbb{D}}}.$$

We note that PCGLS terminated under 150 iterations since it achieved a sufficient reduction in the normal residual, unlike CGLS. Residual plots show that both CGLS and PCGLS are able to reduce the residual down to noise level, $-0.7 \approx \log(20\%)$, though PCGLS was able to achieve this in less than 5 iterations while CGLS took around 15 iterations. Fig. 7 plots the log of the relative norm of the inversion error versus CG iteration index k, i.e.,

$$\log \frac{\|\mathbf{w}_{\text{true}} - \mathbf{w}^{(k)}\|}{\|\mathbf{w}_{\text{true}}\|},$$

with respect to the L^2 norm, and for PCGLS the weighted norm as well. Inversion error plots show that both CGLS and PCGLS achieve a minimal inversion error of less than 2.5% under 20 and 5 iterations respectively. We also observed that the inversion error begins to increase, while the residual plateaus, for latter CG iterations which is illustrative of the semi-convergence (overfitting) behavior of CG when applied to ill-conditioned inverse problems with noisy data. Interestingly enough, preconditioning seems to accelerate both the reduction in residual and inversion error as well as the overfitting. Lastly, for completeness, we plot the true and estimated multipole coefficients in Fig. 8. We mention that the plotted estimates $\mathbf{w}^{(k)}$ correspond to $k = 32$ and $k = 4$ for CGLS and PCGLS respectively, where k was chosen as to yield the minimal inversion error.

Test 2: Estimating dipole source

In the second test case we invert for a dipole source,

$$f(\mathbf{x}, t) = w_1(t)\frac{\partial}{\partial x_2}\delta(\mathbf{x} - \mathbf{x}^*).$$

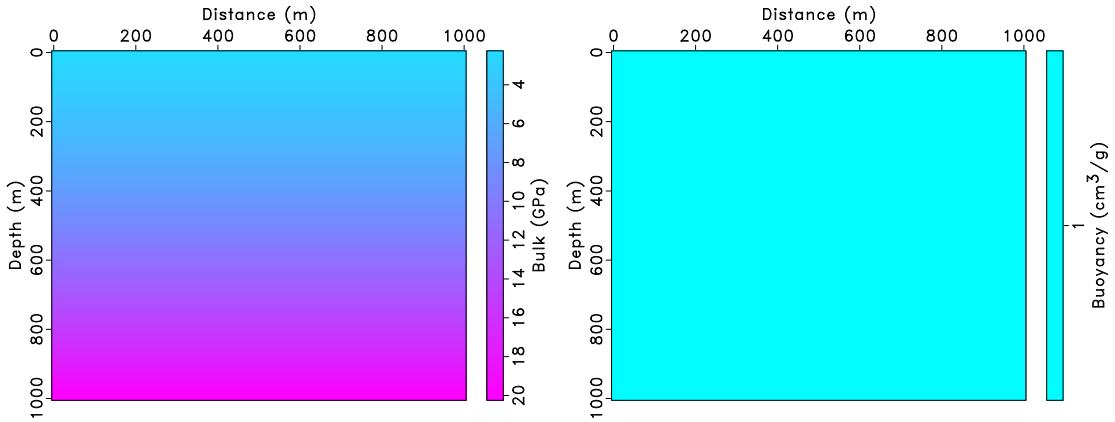
The corresponding (discrete) weight operator $\mathbf{L}$ is given by the fractional 3/2-derivative,

$$\mathbf{L} = \mathbf{D}_t^{3/2},$$

with inverse given by the fractional 3/2-integral,

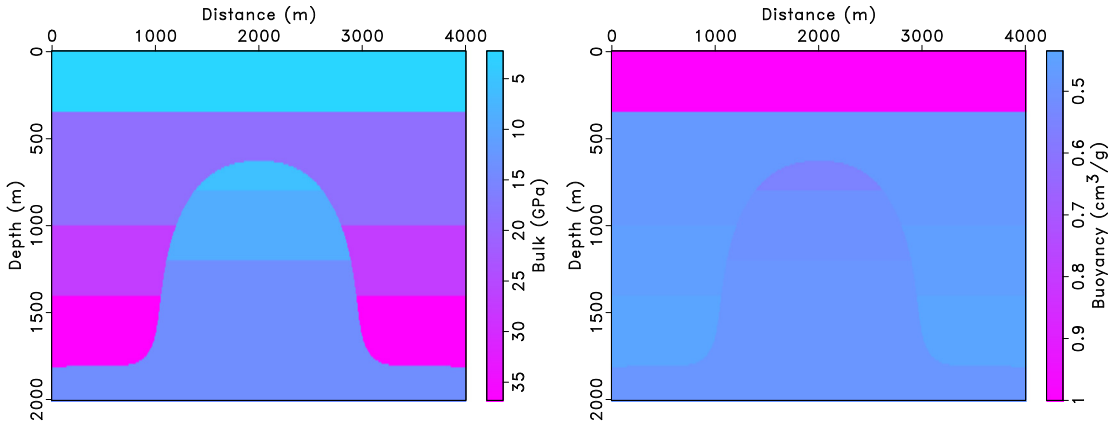
$$\mathbf{L}^{-1} = \mathbf{D}_t^{-3/2}.$$

Fig. 5b plots the noise-free data; noise level of $\eta = 20\%$ is added to observed data. Note the anisotropy observed in the data, specifically on the horizontal line of receivers (first 100 receivers), consistent with the fact that given source is a dipole in the horizontal direction.



(a) Bulk modulus. (b) Buoyancy.

Fig. 3. Linear wave speed model used in test 4.



(a) Bulk modulus. (b) Buoyancy.

Fig. 4. Dome model used in test 5.

Again, both CGLS and PCGLS manage to reduce the residual to noise level, though it took 150 iterations to achieve this without preconditioning (roughly) and under 5 iterations with preconditioning; see Fig. 9. Moreover, PCGLS was able to recover very accurate multipole coefficients, with inversion error less that 3% in under 10 iterations, while CGLS could not achieve an inversion error of less that 30% under 150 iterations; see Figs. 10 and 11.

Test 3: Estimating multipole series in constant medium

In test three we invert for $\mathbf{w} = [w_1, w_2, w_3]^T$ associated with the following first-order multipole series,

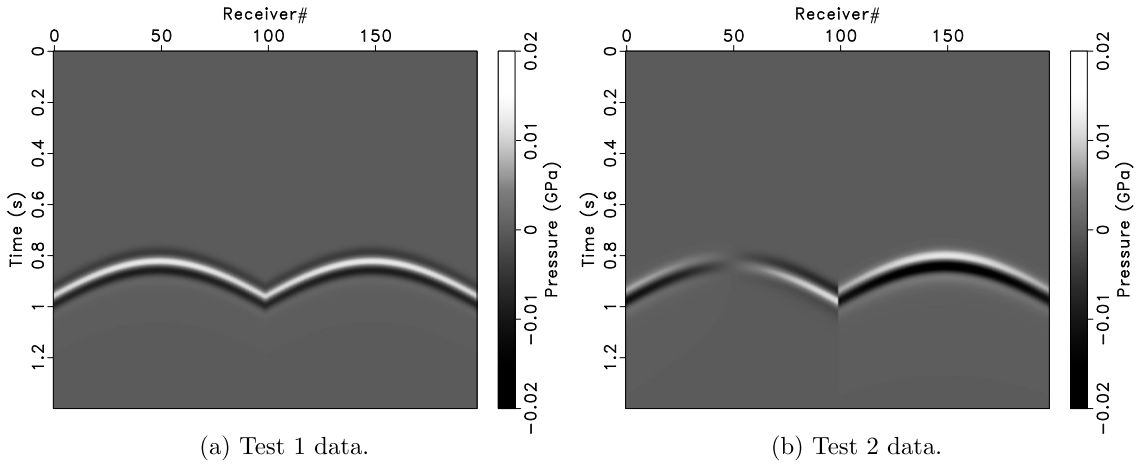
$$f(\mathbf{x}, t) = w_1(t)\delta(\mathbf{x} - \mathbf{x}^*) + w_2(t)\frac{\partial}{\partial x_1}\delta(\mathbf{x} - \mathbf{x}^*) + w_3(t)\frac{\partial}{\partial x_2}\delta(\mathbf{x} - \mathbf{x}^*). \tag{22}$$

The true multipole coefficients were scaled as to yield similar amplitudes in the contribution of each multipole term to the observed data. The (discrete) weight operator $\mathbf{L}$ for this test is given by

$$\mathbf{L} = \begin{bmatrix} \mathbf{D}_t^{1/2} & & \\ & \frac{1}{c}\mathbf{D}_t^{3/2} & \\ & & \frac{1}{c}\mathbf{D}_t^{3/2} \end{bmatrix}$$

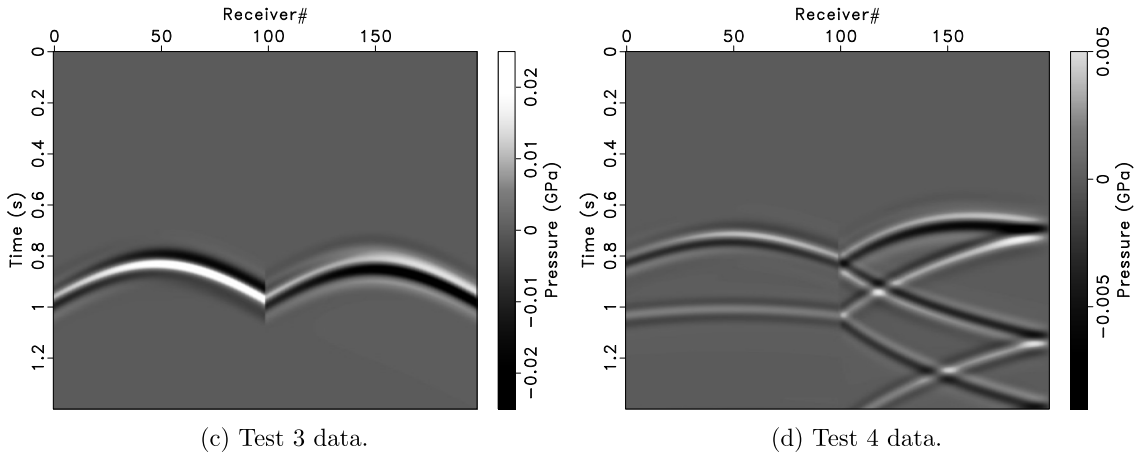
with inverse

$$\mathbf{L}^{-1} = \begin{bmatrix} \mathbf{D}_t^{-1/2} & & \\ & c\mathbf{D}_t^{-3/2} & \\ & & c\mathbf{D}_t^{-3/2} \end{bmatrix},$$



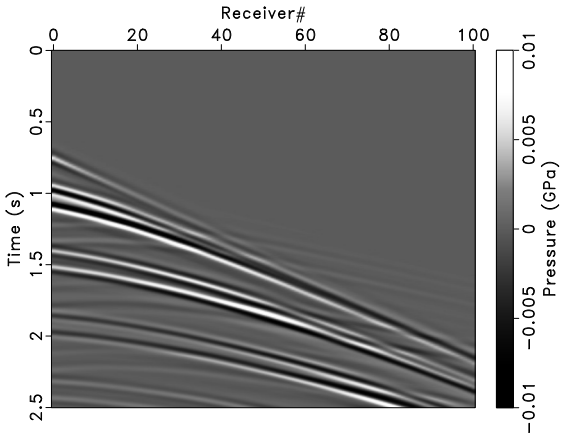
(a) Test 1 data.

(b) Test 2 data.



(c) Test 3 data.

(d) Test 4 data.



(e) Test 5 data.

Fig. 5. Observed (noise-free) data in the form of time traces of pressure field at receiver locations.

where $c = 1.5$ km/s. Noise-free data is shown in Fig. 5c. In this test we consider observed data polluted by two noise levels, $\eta = 20\%$ and $\eta = 1\%$, in order to observe the effect of noise on the inversions.

Residual plots, Fig. 12, demonstrate the ability of PCGLS to reduce the residual to noise level in under 20 iterations, while after 150 iterations CGLS struggled to reduce the residual down to 30%. Moreover, PCGLS was able to produce multipole coefficient estimates with inversion errors of roughly 6% and 2% given 20% and 1% noise levels respectively, again under 20 iterations; Fig. 13. On the other hand, CGLS was only able to achieve inversion errors of roughly 60%. We plot the best estimated multipole coefficients, along with the true solution, for the two noise levels in Figs. 14 and 15.

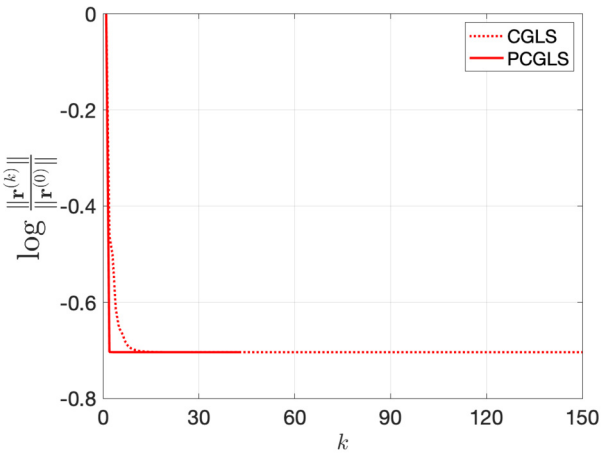


Fig. 6. Evolution of residual (data misfit) at each CG step k for test 1.

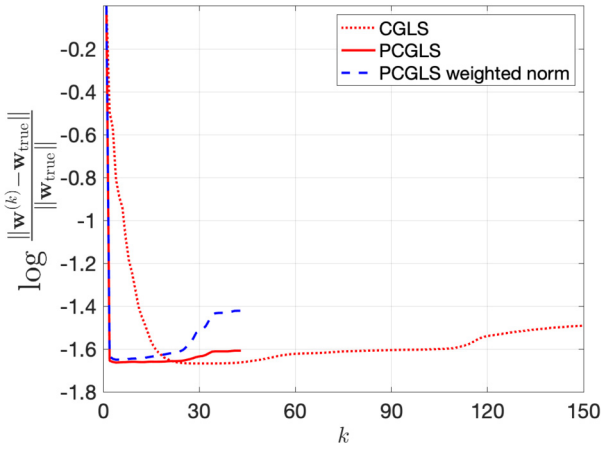


Fig. 7. Evolution of inversion error at each CG step k for test 1, using L^2 norm $\|\cdot\|_{\mathbb{W}}$ and weighted norm $\|\cdot\|_{\mathbb{W}'}$.

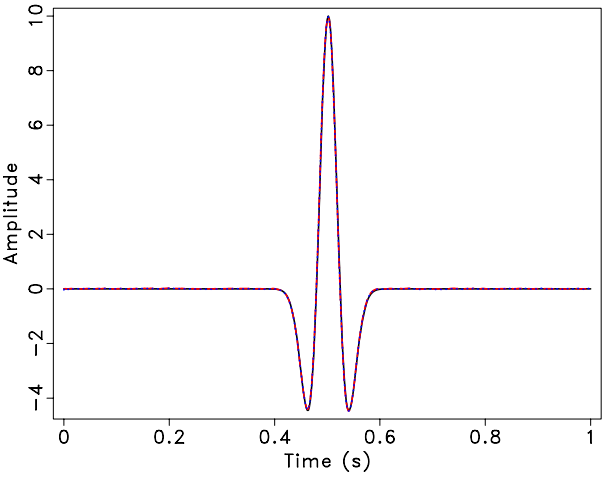


Fig. 8. True solution $\mathbf{w}_{\text{true}}$ (solid) plotted against best estimates $\mathbf{w}^{(k)}$ for test 1; $k = 32$ for CGLS (dotted) and $k = 4$ for PCGLS (dashed).

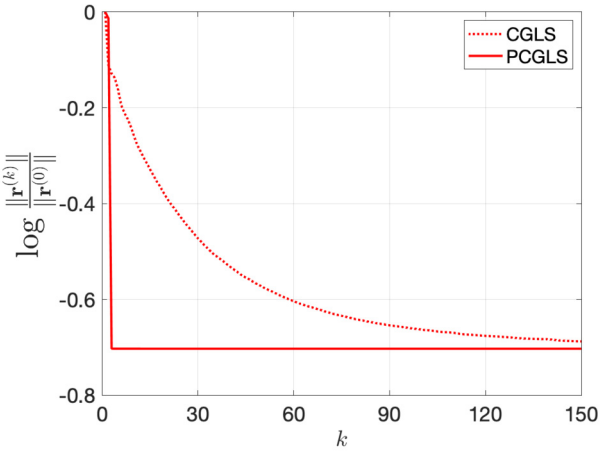


Fig. 9. Evolution of residual (data misfit) at each CG step k for test 2.

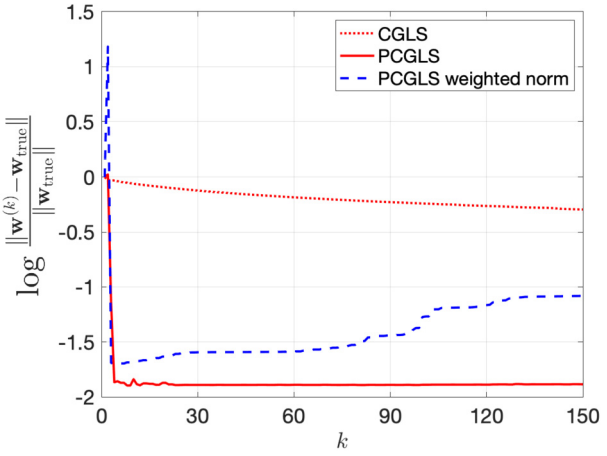


Fig. 10. Evolution of inversion error at each CG step k for test 2, using L^2 norm $\|\cdot\|_{\mathbb{W}}$ and weighted norm $\|\cdot\|_{\mathbb{W}'}$.

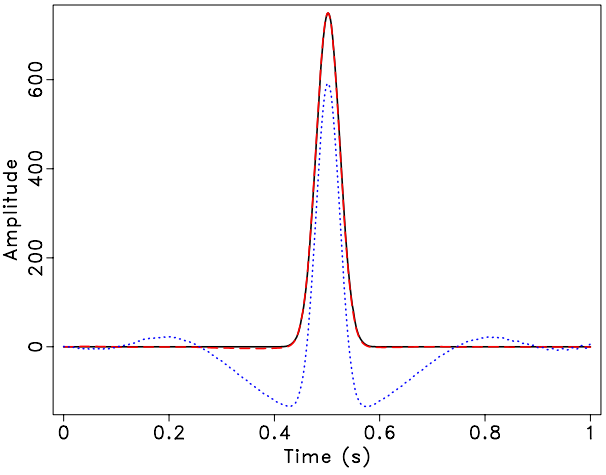
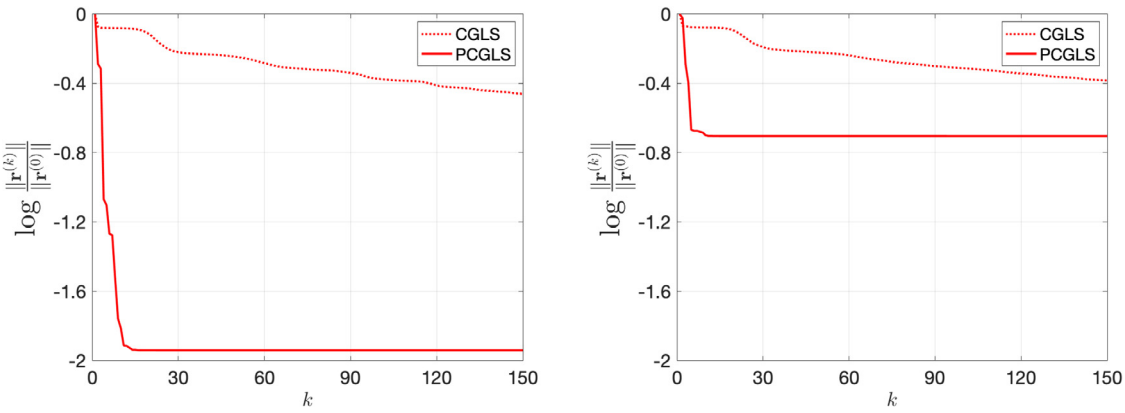
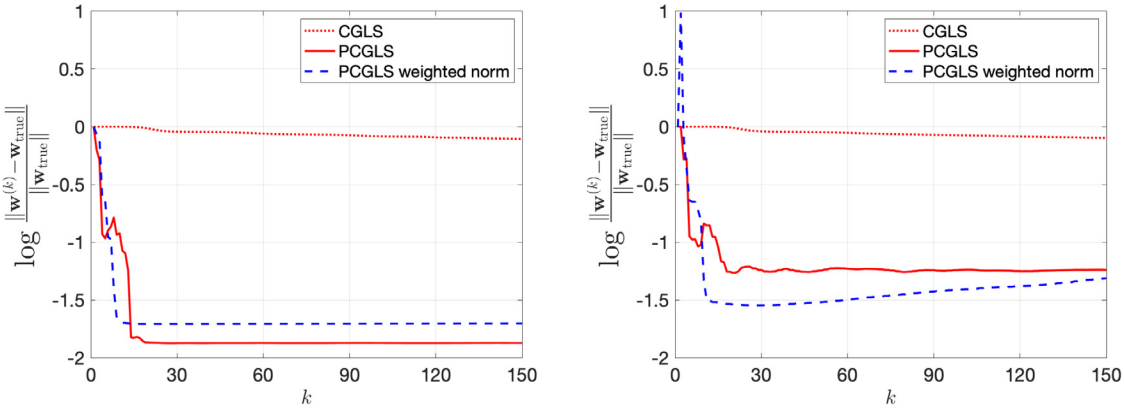


Fig. 11. True solution $\mathbf{w}_{\text{true}}$ (solid) plotted against best estimates $\mathbf{w}^{(k)}$ for test 2; $k = 150$ for CGLS (dotted) and $k = 9$ for PCGLS (dashed).



(a) 1% noise level case (b) 20% noise level case

Fig. 12. Evolution of residual (data misfit) at each CG step k for test 3.



(a) 1% noise level case (b) 20% noise level case

Fig. 13. Evolution of inversion error at each CG step k for test 3, using L^2 norm $\|\cdot\|_{\mathbb{W}}$ and weighted norm $\|\cdot\|_{\mathbb{W}'}$.

Test 4: Estimating multipole series in inhomogeneous medium

Recall that the weight operator, **L**, associated with our preconditioning scheme for multipole source inversion requires scaling by some wave speed c. In the case of a constant medium, as in the previous numerical tests, c is simply taken to be the constant background wave speed. This brings up the following question: How should one construct **L** for inhomogeneous media? It turns out that a simple and effective strategy, even in the presence of media with varying wave speed, is to take c to be a constant value in the ballpark of the background wave speed. The following test shows the effectiveness of this strategy for different values of c in an inhomogeneous medium.

In this test we estimate the same first-order multipole source as in test 3, see equation (22), though in a 2D inhomogeneous medium with wave speed linearly varying in depth from 1.5 km/s to 4.5 km/s and a constant density (or buoyancy), see Fig. 3. The physical domain again consists of $\Omega = [0, 1] \times [0, 1]$ in units of kilometers. PML layers are added to the $x_1 = 0$ and $x_1 = 1$ km boundaries while free surface boundary conditions are applied to the remaining boundaries. Reflective boundaries were added to test the effectiveness of our preconditioner under further inhomogeneities in the medium. The same source-receiver configuration used in the previous tests is also used here. Fig. 5d shows the noise-free data for this test; a noise level of $\eta = 20\%$ was added to generate the observed data. Unlike in the previous tests, observed data for test 4 now includes multiple reflections due to the reflective boundary conditions.

Given that we are inverting for the same multipole source as in the previous test, it follows that **L**, and thus the associated preconditioner, will be of the same form for test 4. What will be different here is the choice of c in **L**. We pick three different values of c: the minimum, average and maximum wave speeds: $c = 1.5, 3.0, 4.5$ in units of km/s. Fig. 16 shows that PCLS, with any of the choices of c, is able to accelerate CG to reduce the residual down to noise level in under 60 iterations, while CGLS was barely able to reduce the residual down to 40% in 150 iterations. Moreover, PCGLS recovers substantially more accurate multipole coefficients, with the best results obtained for $c = 4.5$ km/s; see Fig. 17. (See also Fig. 18.)

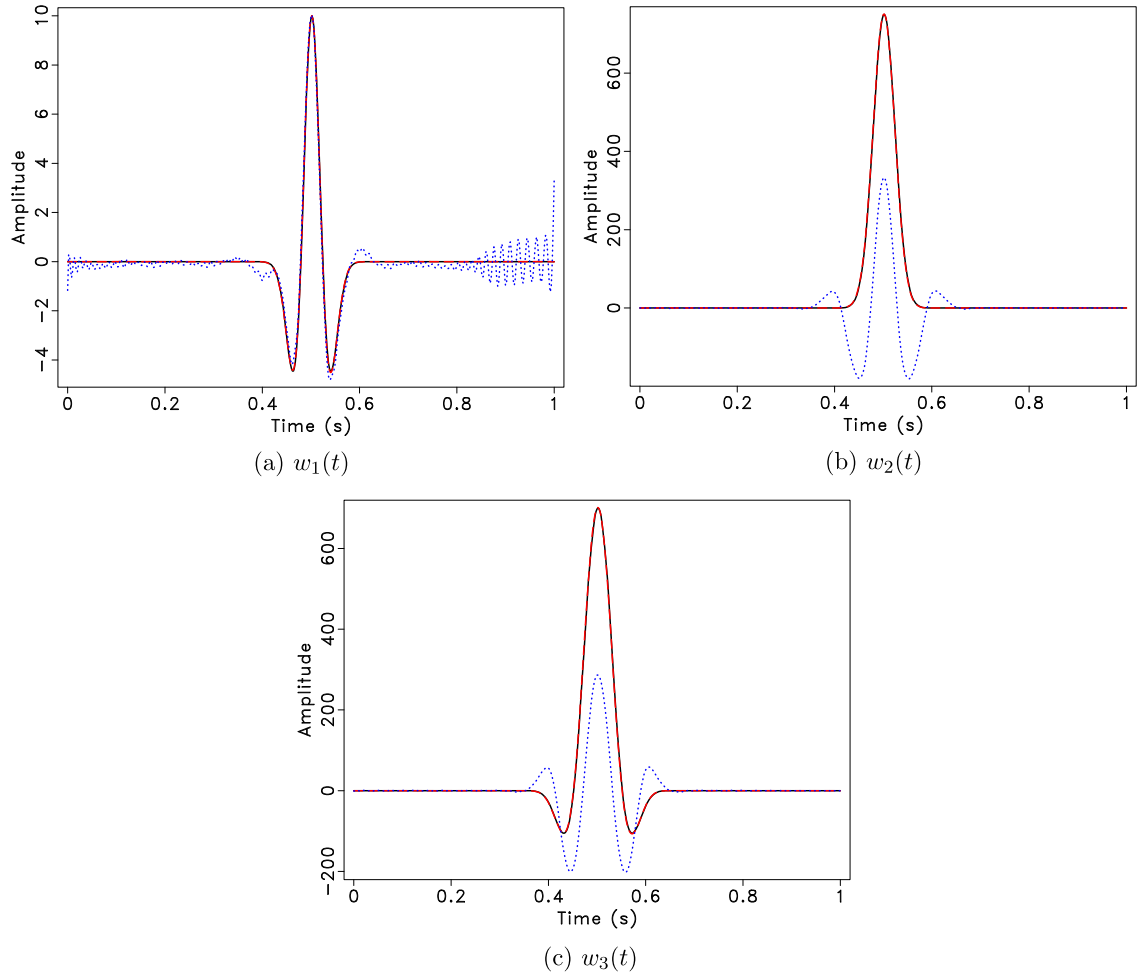


Fig. 14. True solution $\mathbf{w}_{\text{true}}$ (solid) plotted against best estimates $\mathbf{w}^{(k)}$ for test 3 with 1% noise level; $k = 150$ for CGLS (dotted) and $k = 25$ for PCGLS (dashed).

Test 5: Estimating multipole series from synthetic reflectivity data

A fifth and final test is conducted on a more complex data set, namely, reflectivity data resulting from a dome model; see Fig. 4 for plots of medium parameters. The physical domain consists of $\Omega = [0, 2] \times [0, 4]$ in units of kilometers. PML boundaries are added at the $x_2 = 0$, $x_2 = 4$ km, and $x_1 = 2$ km edges of the domain while a free surface boundary condition is applied at the $x_1 = 0$ boundary to simulate the interface between ocean water and air. Source and receivers are located as to simulate a marine seismic reflection survey: source is located at $\mathbf{x}^* = (100, 1000)$ and a single line of receivers are placed on $(10, 1400 : 10 : 3400)$, both in units of meters. In this test we invert for the same multipole source used in tests 3 and 4. Moreover, we also use the same weight operator $\mathbf{L}$ as in the previous test, though with $c = 3.0$ km/s since it produced the best results in practice for this test. Observed (noise-free) data is shown in Fig. 5e; noise level of $\eta = 20\%$ is used for this test.

The source-receiver configuration for test 5, unlike previous tests, is not optimized to capture the entirety of source anisotropy. Again, only a single line of receivers, parallel to the horizontal/x_2 axis, is placed on the right of the source which will fail to capture much of the source anisotropy due to either of the dipole terms. In anticipation of any ill-posedness associated with poor data-coverage for this test we consider the following regularized multipole source FWI problem, i.e.,

$$\min_{\mathbf{w}\in\mathbb{W}} \frac{1}{2}\|\mathbf{F}\mathbf{w} - \mathbf{d}_{\text{obs}}\|^2_{\mathbb{D}} + \frac{\lambda}{2}\|\mathbf{w}\|^2_{\mathbb{W}'}. \tag{23}$$

Note that the regularization term includes the weighted norm of $\mathbf{w}$, which is equivalent to the L^2 norm or $\mathbf{L}\mathbf{w}$. Thus, we solve the regularized FWI problem above by solving the modified normal equations,

$$\left(\mathbf{F}^T\mathbf{F} + \lambda\mathbf{L}^T\mathbf{L}\right)\mathbf{w} = \mathbf{F}^T\mathbf{d}_{\text{obs}}, \tag{24}$$

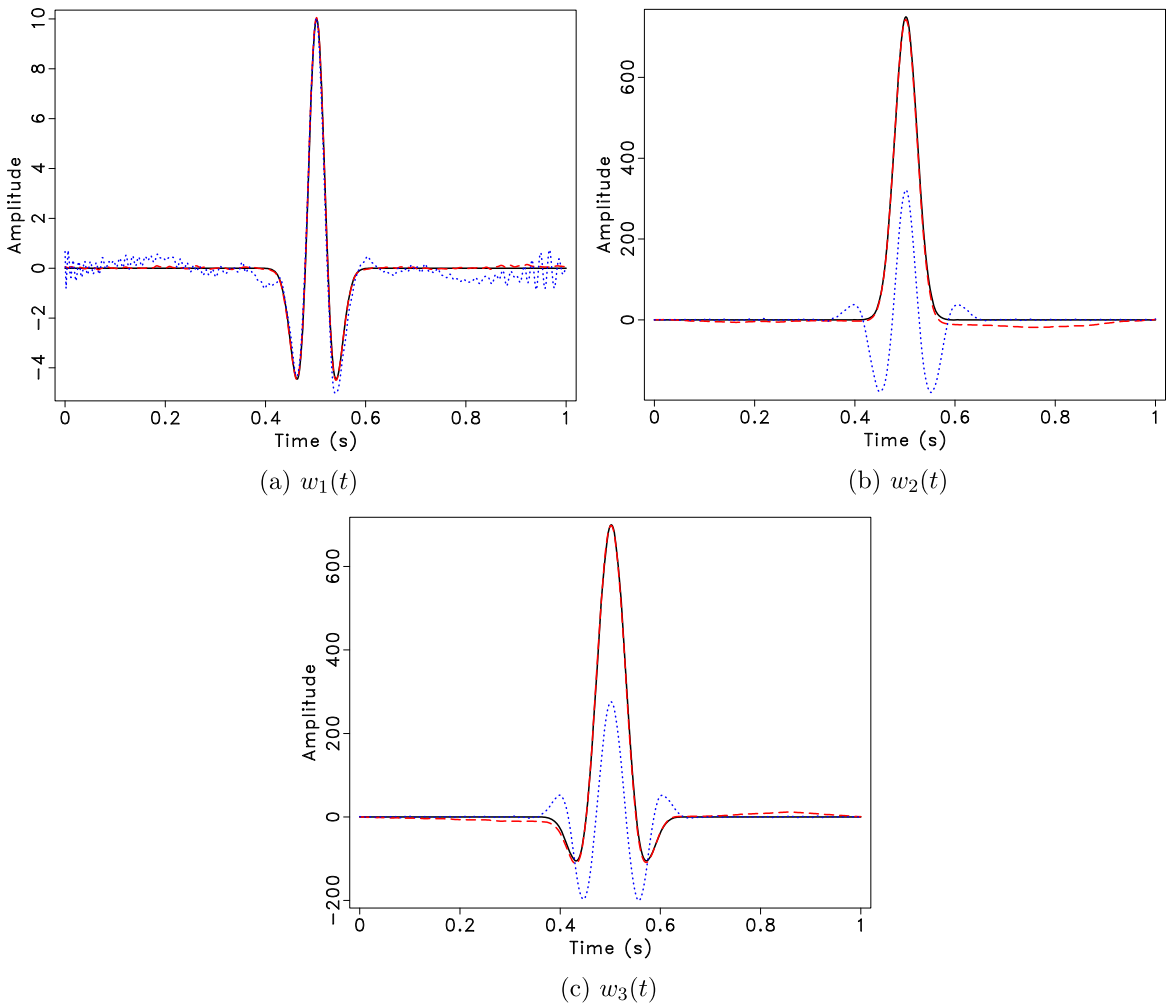


Fig. 15. True solution $\mathbf{w}_{\text{true}}$ (solid) plotted against best estimates $\mathbf{w}^{(k)}$ for test 3 with 20% noise level; $k = 150$ for CGLS (dotted) and $k = 30$ for PCGLS (dashed).

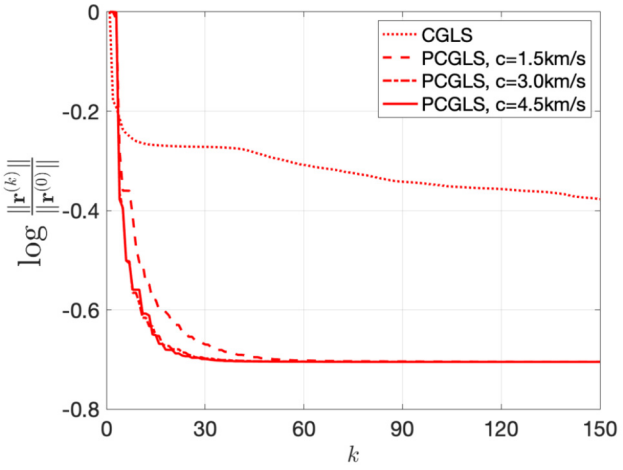
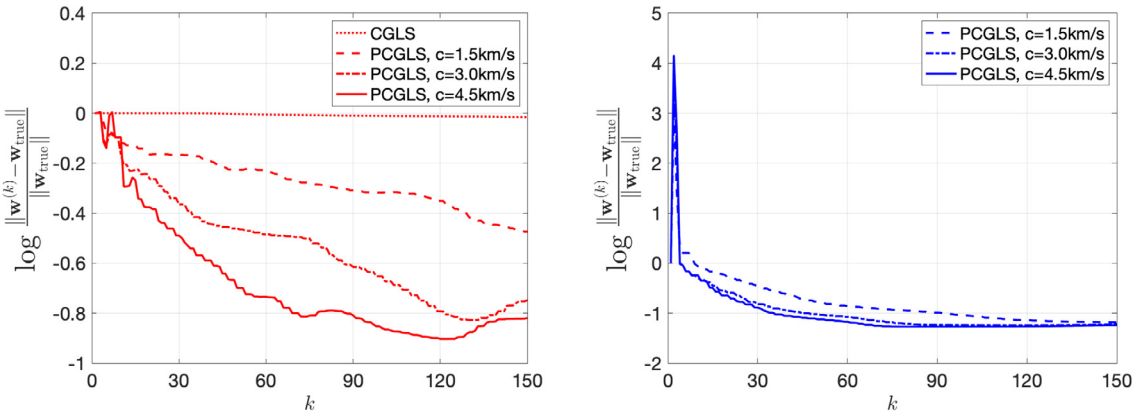


Fig. 16. Evolution of residual (data misfit) at each CG step k for test 4.



(a) Using L^2 norm $\|\cdot\|_{\mathbb{W}}$ (b) Using weighted norm $\|\cdot\|_{\mathbb{W}'}$

Fig. 17. Evolution of inversion error at each CG step k for test 4, using L^2 norm $\|\cdot\|_{\mathbb{W}}$ and weighted norm $\|\cdot\|_{\mathbb{W}'}$.

where λ is the damping constant, also referred to as the regularization parameter. We used a variant of PCGLS for solving the regularized normal equations (24) which we refer to as PCGLS+Reg, essentially the preconditioned version of algorithm 3 discussed in Frommer and Maass [7]. Moreover, we do not show any results involving CGLS+Reg, the non-preconditioned version of PCGLS+Reg, since regularization was shown to have no significant impact on the inversions given how slow CGLS is without preconditioning. We avoid the difficulty associated with choosing regularization parameter λ by simply choosing it in a trial and error manner as to yield the best results, $\lambda = 5 \times 10^{-8}$.

Residual plots, Fig. 19, show that preconditioning is able to accelerate the reduction of data misfit substantially, close to noise level. As suspected, source inversions are not as accurate as in the previous test cases, though preconditioning makes considerably more progress than CGLS in reducing the inversion error; see inversion error plots 20. In Fig. 20b, the vertical axis is cutoff in order to better visualize the behavior of the inversion error with respect to the weighted norm at latter CG iterations. Inversion error plots also demonstrate the effects of regularization: inversion curves plateau after $k = 60$ roughly at the minimal value of 63% in the L^2 norm, thought at the cost of a slightly larger inversion error in the weighted norm. (See also Fig. 21.)

In the context of marine seismic surveys, and other applications where characterizing the seismic source is secondary to seismic imaging/inversion, the source inverse problem constitutes a subproblem in a joint medium-source inversion approach. In this context, it suffices to estimate a source that explains the observed data well enough without overfitting. For that reason we consider source estimates given by PCGLS and PCGLS+Reg for this test case satisfactory. Fig. 22 compares the observed data along with the data misfit for the different algorithms using the same gray scale. The data misfits for preconditioned results are some what visible, though a marked improvement over CGLS. In Fig. 23, we compare a windowed single trace of the observed data with the predicted data from CGLS and PCGLS; PCGLS+Reg is not included in this plot for the sake of visual clarity and since it produced a similar result.

8. Discussions

Overall, preconditioning resulted in the acceleration of CG iterates for all numerical tests. In some cases, reduction of data misfit down to noise level and accurate source inversions under 150 iterations was only achieved with preconditioning. We also note that the effects of preconditioning were more dramatic for the more complicated source. Though preconditioning scheme is based semi-heuristically on the analytical solutions to the wave equation in an unbounded and constant medium, we showed their efficacy in test cases with inhomogeneous media as well.

So far our comparisons between CGLS and PCGLS focus on iteration counts, which is an incomplete picture given that PCGLS is more computationally expensive per iteration. We take this moment to talk about computational cost per CG iteration, with and without preconditioning, as a proxy to the overall computational cost of each iterative method. In particular, we focus on counting the number of forward (or adjoint) map actions as well as the action of the inverse of the weight operator $\mathbf{L}$ (or its inverse adjoint), since these are the most computationally expensive operations at each iteration. We first discuss CGLS, which essentially requires one $\mathbf{F}$ and one $\mathbf{F}^T$ action per iteration. PCGLS also requires the same number of forward and adjoint map actions per iteration, as well as an added $\mathbf{L}^{-1}$ and $\mathbf{L}^{-T}$ action, at least when using a split preconditioner implementation as done here. We note that $\mathbf{L}^{-1}$ is essentially an $(MK_w) \times (MK_w)$ triangular matrices, while on the other hand $\mathbf{F}$ is an $(NK_d) \times (MK_w)$ and block Toeplitz matrix. In the numerical tests carried out here we have that $M = 3$ at most while $N = 200$, i.e., number of multipole coefficients is at least two orders of magnitude smaller than the number of receivers. Thus we can argue that PCGLS is only negligibly more expensive than CGLS per iteration.

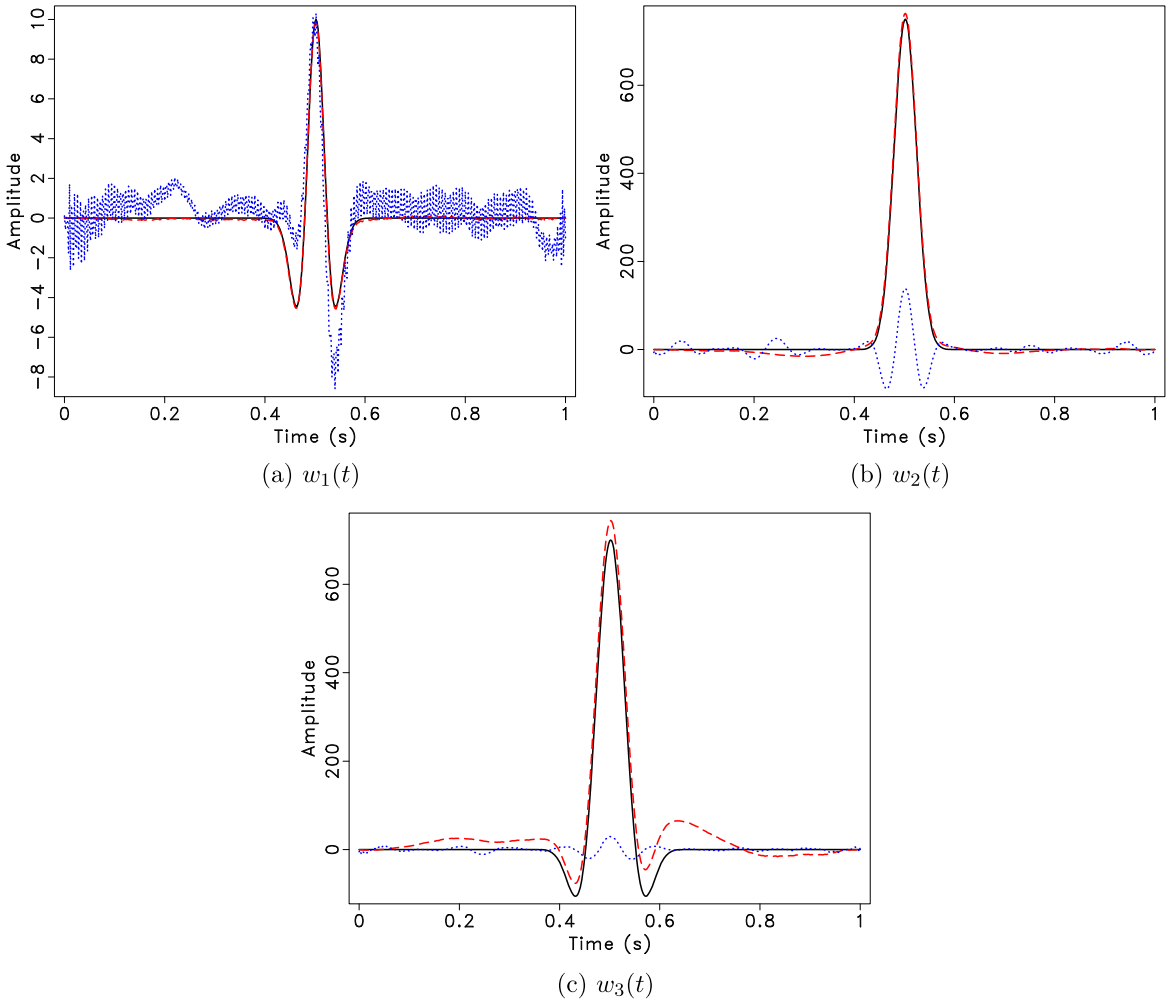


Fig. 18. True solution $\mathbf{w}_{\text{true}}$ (solid) plotted against best estimates $\mathbf{w}^{(k)}$ for test 4; $k = 150$ for CGLS (dotted) and $k = 120$ for PCGLS (dashed) with $c = 4.5$ km/s.

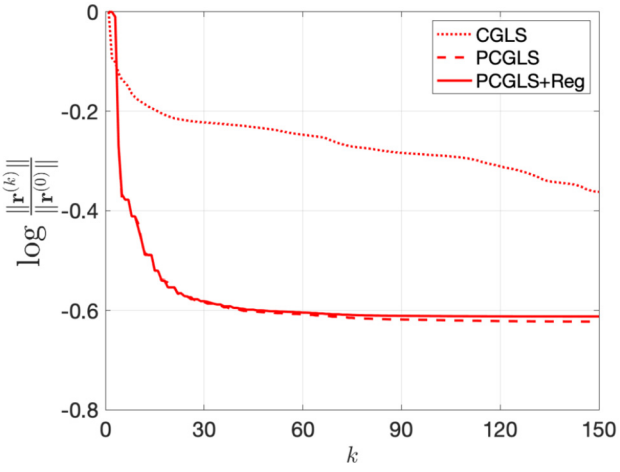
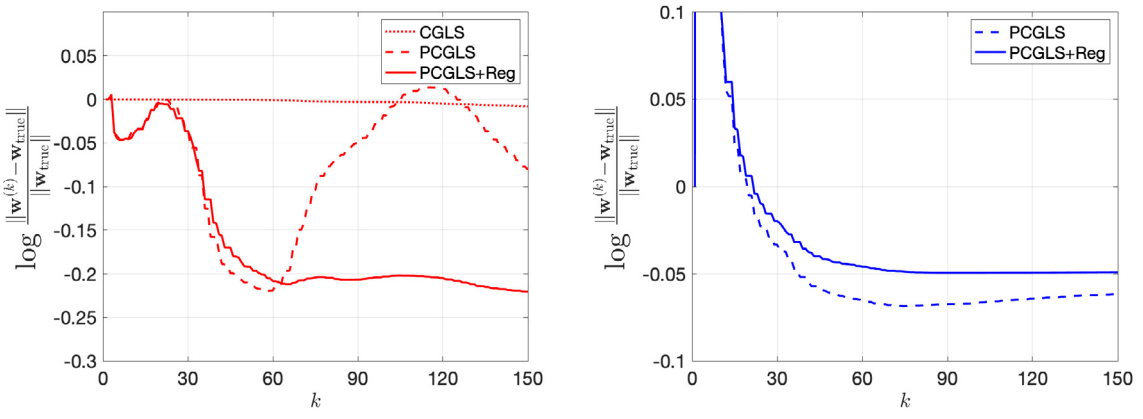
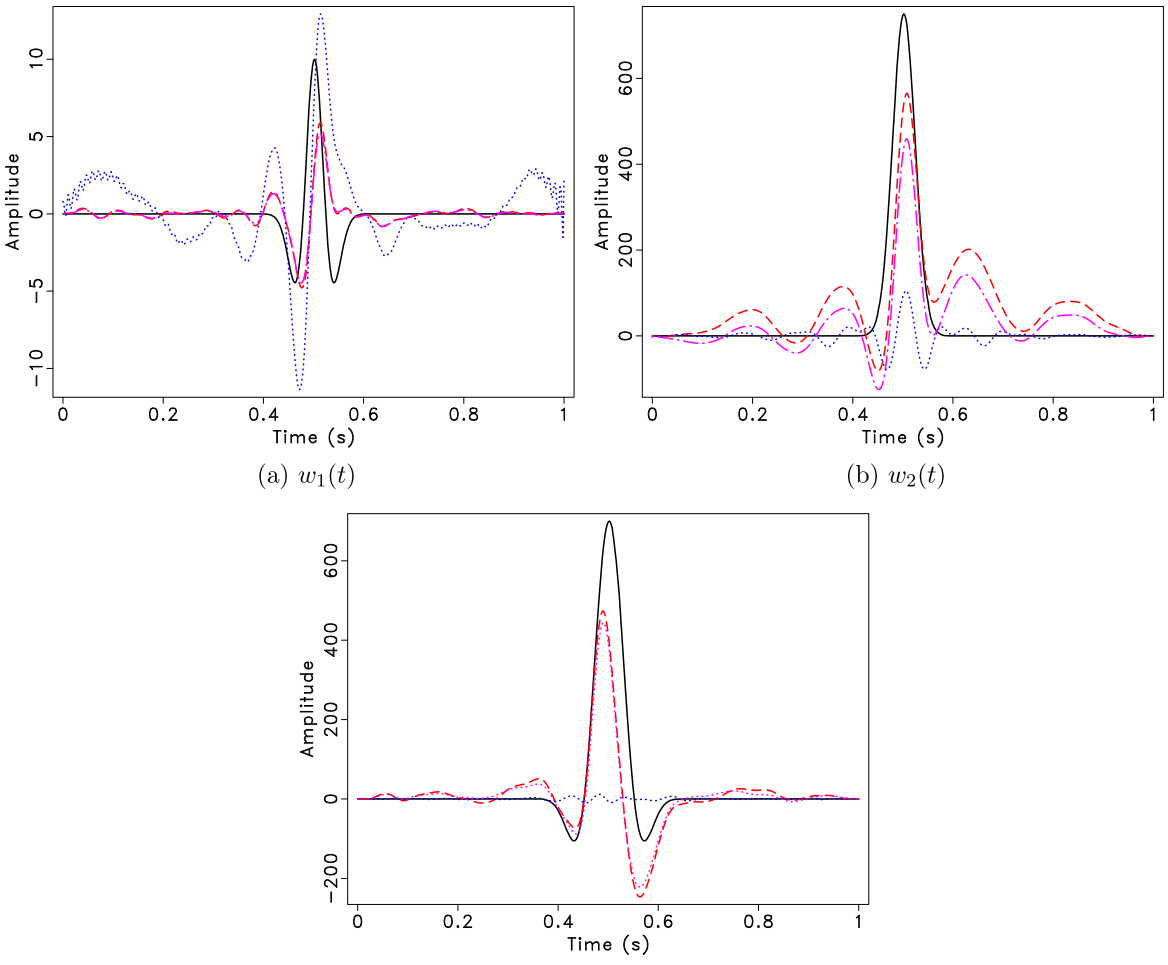


Fig. 19. Evolution of residual (data misfit) at each CG step k for test 5.



(a) Using L^2 norm $\|\cdot\|_{\mathbb{W}}$ (b) Using weighted norm $\|\cdot\|_{\mathbb{W}'}$

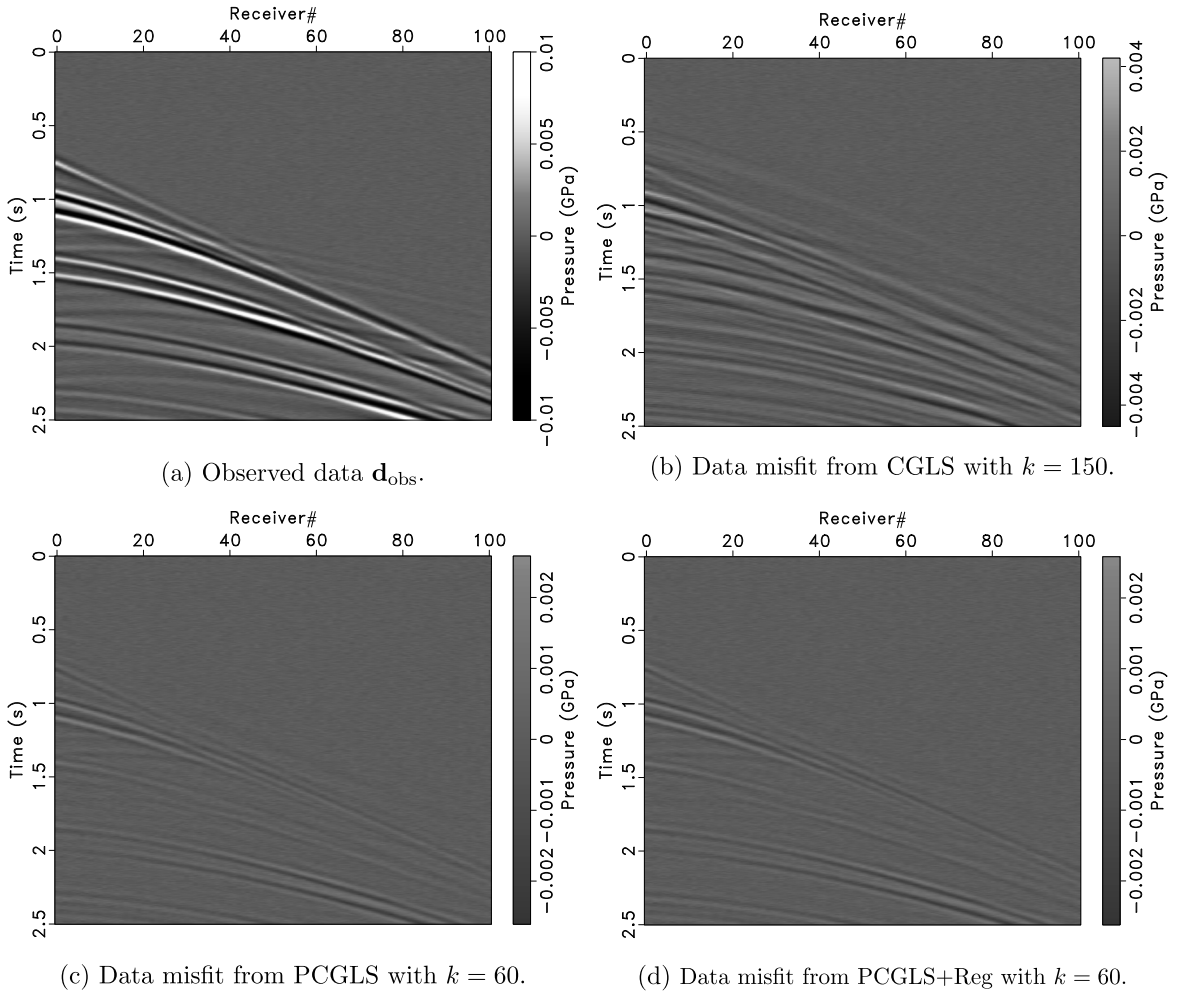
Fig. 20. Evolution of inversion error at each CG step k for test 5, using L^2 norm $\|\cdot\|_{\mathbb{W}}$ and weighted norm $\|\cdot\|_{\mathbb{W}'}$.



(a) $w_1(t)$ (b) $w_2(t)$

(c) $w_3(t)$

Fig. 21. True solution $\mathbf{w}_{\text{true}}$ (solid) plotted against best estimates $\mathbf{w}^{(k)}$ for test 5; $k = 150$ for CGLS (dotted), and $k = 60$ for PCGLS (dashed) and PCGLS+Reg (dash-dotted).



(a) Observed data $\mathbf{d}_{\text{obs}}$.

(b) Data misfit from CGLS with $k = 150$.

(c) Data misfit from PCGLS with $k = 60$.

(d) Data misfit from PCGLS+Reg with $k = 60$.

Fig. 22. Comparing observed data $\mathbf{d}_{obs}$ with data misfits $\mathbf{d}_{\text{obs}} - \mathbf{F}\mathbf{w}^{(k)}$ for test 5.

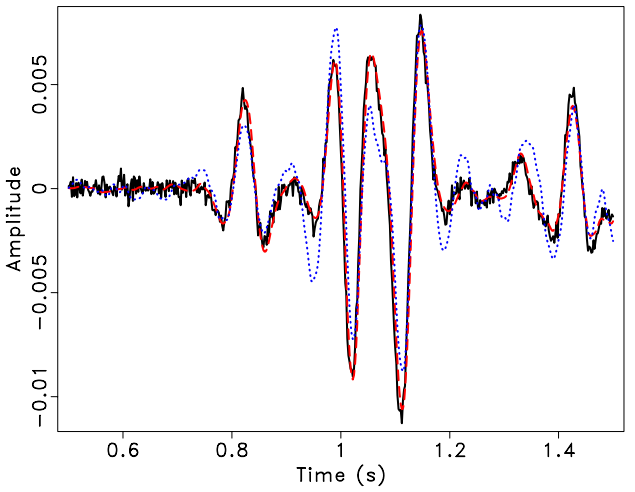


Fig. 23. Comparing observed data $\mathbf{d}_{\text{obs}}$ (solid) and predicted data $\mathbf{F}\mathbf{w}^{(k)}$ at receiver 5; $k = 150$ for CGLS (dotted), and $k = 60$ for PCGLS (dashed).

$$\text{relative cost per iteration of PCGLS} \sim \frac{NM + M^2}{NM} = 1 + \frac{M}{N}.$$

In the extreme case where $N \approx M$, we have that PCGLS is twice as expensive as CGLS per iteration. In either case, the observed acceleration of CG iterates more than makes up for the computational costs associated with preconditioning. We should also mention that PCGLS+Reg requires the same costs as PCGLS with two added actions of $\mathbf{L}$ per iteration,

$$\text{relative cost per iteration of PCGLS+Reg} \sim 1 + \frac{2M}{N},$$

which again, is more than compensated by the acceleration achieved with preconditioning.

Lastly, we discuss our choice of stopping criterion which was meant to allow us to focus on the effects of our preconditioning scheme. In practice, an optimal CG stopping criterion may incorporate information about the noise in the data, and should take advantage of the regularization effects of early termination. Consider the CGLS algorithm. It is well known that the residual monotonically decreases with iteration index while the L^2-norm of the solution monotonically increases, if a zero initial guess is used and under exact arithmetic. In other words, an early termination of CGLS effectively provides Tikhonov regularization where the terminating iteration count plays the role of regularization parameter. The L-curve method consists of plotting the residual and solution norm at varying regularization parameter values yielding a characteristic L shaped curve, and selecting the regularization parameter closest to the corner of the L. The idea behind the L-curve criterion is that the corner point produces a regularized solution that is a balance between overfitting (smaller residual and larger solution norm) and over-regularizing (smaller solution norm and larger residual). Given that the residual and solution norm are rather accessible during CGLS solution update makes the L-curve stopping criterion an attractive choice; we refer to Hansen [10], chapter 6 and 7 for a full discussion of the L-curve and its application as a stopping criterion for CGLS. In the case of PCGLS, with preconditioner of the form $\mathbf{M} = \mathbf{L}^T\mathbf{L}$, it turns out we can relate this algorithm to CGLS via the following standard form transformation, $\tilde{\mathbf{w}} \leftarrow \mathbf{L}\mathbf{w}$. Thus we have

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{F}\mathbf{w} - \mathbf{d}_{\text{obs}}\|_{\mathbb{D}} = \min_{\tilde{\mathbf{w}}} \frac{1}{2}\|\tilde{\mathbf{F}}\tilde{\mathbf{w}} - \mathbf{d}\|_{\mathbb{D}}$$

where $\tilde{\mathbf{F}} = \mathbf{F}\mathbf{L}^{-1}$. It follows that PCGLS produces iterates with monotonically increasing $\|\mathbf{L}\mathbf{w}^{(k)}\|_{\mathbb{W}} = \|\mathbf{w}^{(k)}\|_{\mathbb{W}'}$ which implies that one should monitor the weighted norm of the solution when applying the L-curve stopping criterion. Note that this would be a form of generalized Tikhonov regularization with Tikhonov operator $\mathbf{L}$.

9. Conclusion

In this paper we have presented a framework for inverting seismic sources as series of multipoles, a generalization of moment tensors in earthquake seismology. The source estimation problem reduces to solving normal equations for multipole coefficients. Our main contribution has been to introduce a preconditioner when solving the aforementioned normal equations via CG. This preconditioner consists of (potentially fractional) derivative/integral operators whose order is based semi-heuristically on the analytical solutions of the acoustic wave equation in unbounded constant media with multipole source terms. Numerical experiments conducted show the dramatic acceleration of CG iterates and accuracy of estimated sources when using preconditioning.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to the sponsors of The Rice Inversion Project for their long-term support, and to the Rice Graduate Education for Minorities (RGEM) and XSEDE scholarship programs for their support of M. Bencomo's Ph.D. research. This material is also based upon work supported by the National Science Foundation under Grant No. DMS-1439786 while the author was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Fall 2017 semester.

References

[1] K. Aki, P. Richards, Quantitative Seismology: Theory and Methods, Freeman, 1980.
[2] G. Backus, M. Mulcahy, Moment tensors and other phenomenological descriptions of seismic sources – I. Continuous displacements, Geophys. J. Int. 46 (1976) 341–361.
[3] M.J. Bencomo, W.W. Symes, Discretization of multipole sources in a finite difference setting for wave propagation problems, J. Comput. Phys. 386 (2019) 296–322.

[4] K. Bube, P. Lailly, P.E. Sacks, F. Santosa, W. Symes, Simultaneous determination of source wavelet and velocity profile using impulsive point-source data from a layered fluid, Geophys. J. 95 (1988) 449–462.
[5] J.R. Bunch, Stability of methods for solving Toeplitz systems of equations, SIAM J. Sci. Stat. Comput. 6 (1985) 349–364.
[6] D.W. Eaton, F. Forouhideh, Microseismic moment tensors: the good, the bad and the ugly, CSEG Rec. 35 (2010) 45–49.
[7] A. Frommer, P. Maass, Fast cg-based methods for Tikhonov–Phillips regularization, SIAM J. Sci. Comput. 20 (1999) 1831–1850.
[8] G. Golub, V. Pereyra, The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate, SIAM J. Numer. Anal. 10 (1973) 413–432.
[9] G. Golub, V. Pereyra, Separable nonlinear least squares: the variable projection method and its applications, Inverse Probl. 19 (2003) R1–R26.
[10] P.C. Hansen, Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion, 4th ed., SIAM, 2005.
[11] L. Hörmander, Linear Partial Differential Operators, 3rd ed., Springer-Verlag, 1969.
[12] W. Hu, A. Abubakar, T. Habashy, Application of the nearly perfectly matched layer in acoustic wave modeling, Geophysics 72 (2007) SM169–SM176.
[13] K. Koch, Moment tensor inversion of local earthquake data—I. Investigation of the method and its numerical stability with model calculations, Geophys. J. Int. 106 (1991) 305–319.
[14] M. Landrø, R. Sollie, Source signature determination by inversion, Geophysics 57 (1992) 1633–1640.
[15] R. Lewis, Source-velocity identification for a layered model of reflection seismology, PhD thesis, Department of Mathematical Sciences, Rice University, Houston, TX, USA, 1989.
[16] C. Li, F. Zeng, Numerical Methods for Fractional Calculus, vol. 24, CRC Press, 2015.
[17] M. Li, J. Rickett, A. Abubakar, Application of the variable projection scheme to frequency-domain full-waveform inversion, Geophysics 78 (2013) R249–R257.
[18] R. Madariaga, Dynamics of an expanding circular fault, Bull. Seismol. Soc. Am. 66 (1976) 163–182.
[19] S. Minkoff, W. Symes, Estimating the energy source and reflectivity by seismic inversion, Inverse Probl. 11 (1995) 383–395 (also in Proc. SPIE, 1994).
[20] S.E. Minkoff, W. Symes, Full waveform inversion of marine reflection data in the plane-wave domain, Geophysics 62 (1997) 540–553.
[21] J. Nocedal, S. Wright, Numerical Optimization, Springer-Verlag, 1999.
[22] D. Oldenburg, Multichannel appraisal deconvolution, Geophys. J. Int. 69 (1982) 405–414.
[23] A.D. Padula, W. Symes, S.D. Scott, A software framework for the abstract expression of coordinate-free linear algebra and optimization algorithms, ACM Trans. Math. Softw. 36 (2009) 8.
[24] C.C. Paige, M.A. Saunders, Lsqr: an algorithm for sparse linear equations and sparse least squares, ACM Trans. Math. Softw. 8 (1982) 43–71.
[25] N.A. Petersson, B. Sjögreen, Stable grid refinement and singular source discretization for seismic wave simulations, Commun. Comput. Phys. 8 (2010) 1074–1110.
[26] A. Ramm, Inversion of the backscattering data and a problem of integral geometry, Phys. Lett. A 113 (1985) 172–176.
[27] J. Rickett, The variable projection method for waveform inversion with an unknown source function, in: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2012, pp. 1–5.
[28] E. Robinson, Predictive decomposition of seismic traces, Geophysics 22 (1957) 767–778.
[29] E.A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Goose Pond Press, 1983.
[30] A. Ruhe, P.Å. Wedin, Algorithms for separable nonlinear least squares problems, SIAM Rev. 22 (1980) 318–337.
[31] P.M. Shearer, Introduction to Seismology, Cambridge University Press, 2009.
[32] S.A. Sipkin, Estimation of earthquake source parameters by the inversion of waveform data: synthetic waveforms, Phys. Earth Planet. Inter. 30 (1982) 242–259.
[33] F. Song, M.N. Toksöz, Full-waveform based complete moment tensor inversion and source parameter estimation from downhole microseismic data for hydrofracture monitoring, Geophysics 76 (2011) WC103–WC116.
[34] B.W. Stump, L.R. Johnson, Higher-degree moment tensors – the importance of source finiteness and rupture propagation on seismograms, Geophys. J. Int. 69 (1982) 721–743.
[35] W. Symes, D. Sun, M. Enriquez, From modelling to inversion: designing a well-adapted simulator, Geophys. Prospect. 59 (2011) 814–833, https://doi.org/10.1111/j.1365-2478.2011.00977.x.
[36] W.F. Trench, An algorithm for the inversion of finite Toeplitz matrices, J. Soc. Ind. Appl. Math. 12 (1964) 515–522.
[37] T. Ulrych, Application of homomorphic deconvolution to seismology, Geophysics 36 (1971) 650–660.
[38] J. Waldén, On the approximation of singular source terms in differential equations, Numer. Methods Partial Differ. Equ. 15 (1999) 503–520.
[39] K. Wang, J.R. Krebs, D. Hinkley, A. Baumstein, et al., Simultaneous full-waveform inversion for source wavelet and earth model, in: 87th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2009, pp. 2537–2541.
[40] C. Zhou, G.T. Schuster, S. Hassanzadeh, J.M. Harris, Elastic wave equation traveltime and waveform inversion of crosswell data, Geophysics 62 (1997) 853–868.
[41] A. Ziolkowski, Why don't we measure seismic signatures?, Geophysics 56 (1991) 190–201.