

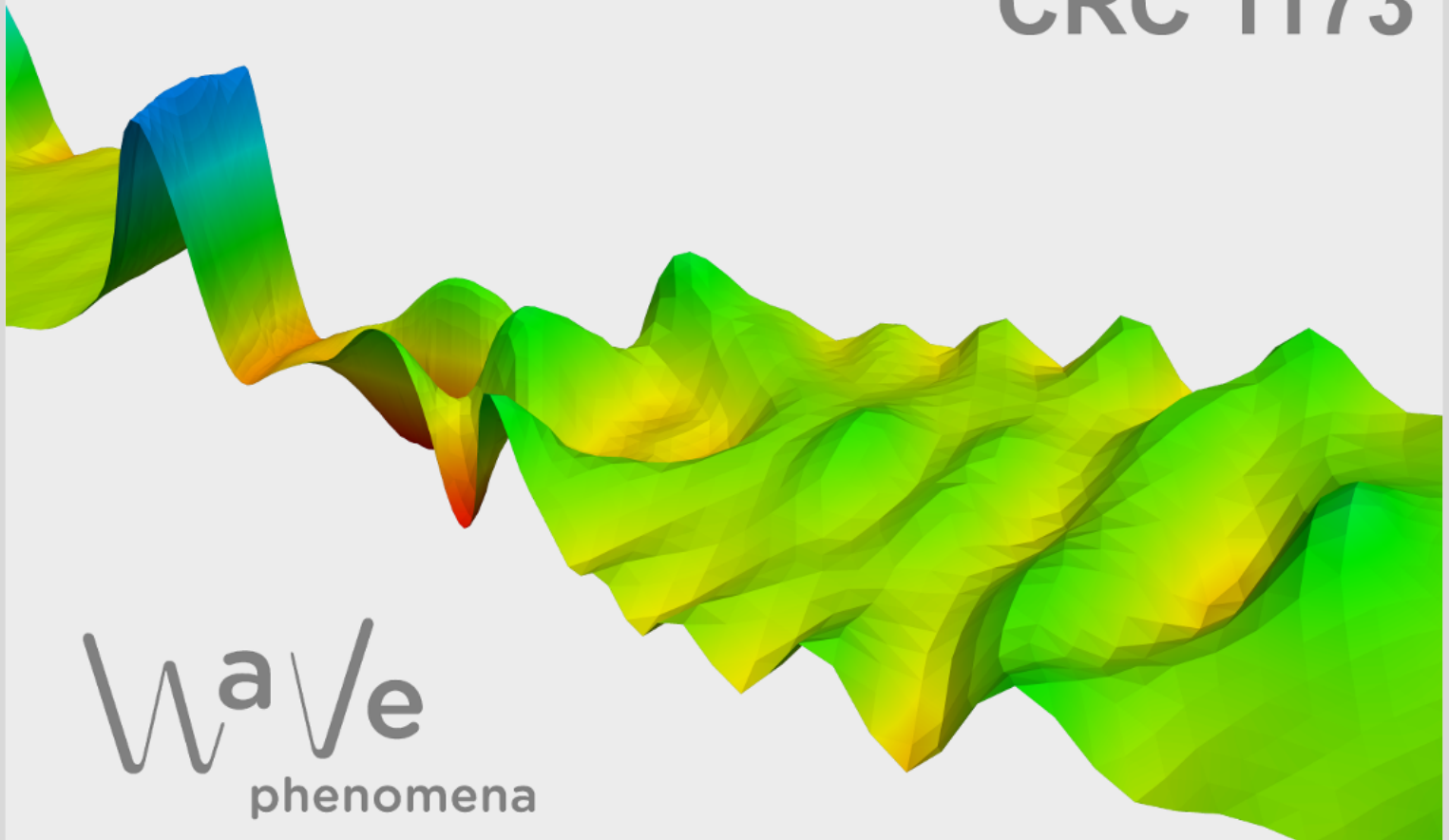
A microlocal prespective on Kirchhoff migration

Andreas Rieder, Kevin Ganster, Eric Todd Quinto

CRC Preprint 2026/56, October 2026

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Wave
phenomena

Participating universities



Funded by



A MICROLOCAL PERSPECTIVE ON KIRCHHOFF MIGRATION

ANDREAS RIEDER¹ , KEVIN GANSTER¹ , AND ERIC TODD QUINTO² 

ABSTRACT. This paper introduces a novel mathematical framework for understanding Kirchhoff migration, a widely used method in seismic imaging for reconstructing subsurface structures from reflection data. By framing Kirchhoff migration operators as generalized Radon transforms and pseudodifferential operators, the study bridges classical seismic processing with tools from microlocal analysis. This perspective expands on, combines, and unifies prior partial results. It enables a detailed examination of how different migration formulas resolve subsurface singularities, such as reflectors and discontinuities, and clarifies the role of acquisition geometry, aperture, and macro velocity models in the quality of the resulting image. Notably, the exposition is largely self-contained and maintains mathematical terminology to a minimum, making the concepts accessible to a broad geophysical audience.

The analysis explains why certain migration operators succeed or fail in preserving true amplitudes or recovering features at specific depths or orientations. It also highlights the conditions under which common migration artifacts arise and how they can be mitigated through operator design. Overall, this approach not only deepens the theoretical understanding of Kirchhoff migration but also provides practical guidance for developing and choosing migration algorithms that offer improved precision and stability, even in complex geological settings, as realistic numerical experiments for the Marmousi model document.

1. INTRODUCTION

The term *Kirchhoff migration* subsumes a variety of linear formulas and algorithms to determine the subsurface pressure wave velocity from reflected wave field measurements, assuming a macro velocity model. Essentially, these formulas can be motivated and derived in two different ways. For instance, the authors of [31, 32] employ a physics-based wave propagation model utilizing ray theory and a simplified earth model, while the authors of [3] rely on a mathematics-based approach. The former approach aligns with the classical geophysical methodology, originating from [20]. The latter approach, on the other hand, emerged from [5] and was subsequently generalized in [2], modifying and extending results of [1], which are based on approximate inversion formulas for generalized Radon transforms (seen as Fourier integral/pseudodifferential operators). The relationship between the constructive technique described in [20] and the mathematics-based approach as used in [3] is explored in [4].

¹DEPARTMENT OF MATHEMATICS, KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT), D-76128 KARLSRUHE, GERMANY

²DEPARTMENT OF MATHEMATICS, TUFTS UNIVERSITY, MEDFORD, MA 02155, USA
E-mail addresses: andreas.rieder@kit.edu, kevin.ganster@kit.edu, todd.quinto@tufts.edu.

Date: October 21, 2026.

Key words and phrases. True-amplitude Kirchhoff migration, seismic imaging, generalized Radon transform.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 258734477 - SFB 1173. The third author's work was partially funded by Simons Foundation Grant MPS-TSM-00708556.

Recently, in [13], we expanded the approach of [1, 34] to interpret these established Kirchhoff migration formulas as general pseudodifferential operators, enabling a precise mathematical analysis. In the present work, we aim to introduce the significance and usefulness of [13] to a broader audience by providing an intuitive interpretation of the results, omitting mathematical details whenever feasible. Nevertheless, we must retain certain mathematical notation and concepts for clarity.

The paper is organized as follows.

In Section 2 we formulate seismic migration as a mathematical inverse problem: we show that all Kirchhoff migration formulas, despite their many variants, share one common structure. They differ only in how the recorded seismic data are filtered, weighted, and summed along diffraction travel-time curves. Readers mainly interested in the practical side will find here our unified description of the classical formulas known from the literature, translated into plain geophysical terms.

The mathematical toolbox — microlocal analysis — is provided in Section 3, in a deliberately informal and self-contained way. Its central question is very intuitive: *Which reflectors can a given migration formula actually image, and how sharply can it image them?* The key concept, the wave front set, describes exactly where in the subsurface a discontinuity lies and in which direction it is oriented — precisely the information a geophysicist needs to judge the quality of a migrated image.

Next, in the central part of the paper (Section 4), we compare the well-known migration formulas of Section 2 using the tools of Section 3 and determine, for each of them, which reflector dips and positions we can reconstruct reliably and which we cannot. We explain why some formulas preserve true amplitudes and others do not, what the acquisition geometry (aperture, offset) contributes, and why certain artifacts appear. Most importantly, in Section 4.5 we recommend simple modifications of the classical formulas that measurably improve image quality — in particular at larger depths and for steeply dipping reflectors.

To put our theory to the test, we choose the well-known Marmousi model, a standard benchmark of realistic complexity (Section 5). Our experiments confirm the theoretical predictions and show how our recommended formulas perform when the idealized assumptions of the theory are violated — as they always are in practice.

Technical background is collected in the appendices: the normalization of the amplitude function, a short introduction to distributions and their Fourier transform, and the details of our efficient, parallelizable implementation of the migration formulas that runs on standard hardware.

Finally, we encourage the reader not to stop after Section 2: the true payoff of the microlocal perspective — our practical guide to choosing and designing migration operators — unfolds in Sections 4 and 5.

2. MIGRATION IN MATHEMATICAL TERMS

The objective of this section is twofold. First, we derive and formulate the task of migration as an inverse tomography problem with a generalized Radon transform, introducing our mathematical notation in the process. Second, we provide the geophysical equivalents of the mathematical terms to make the content of the paper easier to access.

We consider wave propagation in two spatial dimensions¹ and in its most basic form, namely in the acoustic regime with a constant bulk density of 1. Then, the pressure wave field $u =$

¹This is not a principal restriction. It merely simplifies the presentation. For 3D results see [13, Sec. 6].

$u(t; \mathbf{x}, \mathbf{x}_s)$, at $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$ and time $t \geq 0$ satisfies the acoustic wave equation

$$(1) \quad \frac{1}{\nu_p^2} \partial_t^2 u - \Delta_{\mathbf{x}} u = \delta(\mathbf{x} - \mathbf{x}_s) \delta(t), \quad u|_{t=0} = \partial_t u|_{t=0} = 0,$$

where $\nu_p = \nu_p(\mathbf{x})$ is the compression wave speed and wave propagation is initiated by a source located at $\mathbf{x}_s \in \mathbb{R}^2$ and fired at time $t = 0$ (δ denotes the Dirac distribution in the respective variables and $-\Delta_{\mathbf{x}} = -\partial_{x_1}^2 - \partial_{x_2}^2$ is the Laplacian differential operator).

The seismic inverse problem associated with (1) is to recover ν_p in an open set $X \subset \mathbb{R}_+^2 := \mathbb{R} \times [0, \infty)$ from reflected fields $u(t; \mathbf{x}_r, \mathbf{x}_s)$, $(t; \mathbf{x}_r, \mathbf{x}_s) \in [0, T] \times \mathfrak{R} \times \mathfrak{S}$. Here, T is the observation time. The discrete sets $\mathfrak{S}$ and $\mathfrak{R}$ contain the source and receiver positions, respectively. They are subsets of the boundary $\partial\mathbb{R}_+^2 = \mathbb{R} \times \{0\}$ of the propagation medium. The depth coordinate x_2 is positive and points downwards.

The above-described inverse problem is nonlinear. We linearize it by the Born ansatz

$$(2) \quad \frac{1}{\nu_p^2(\mathbf{x})} = \frac{1 + n(\mathbf{x})}{v^2(\mathbf{x})}$$

with a smooth, a priori known reference/macro velocity model $v = v(\mathbf{x})$. The dimensionless correction term n in (2) captures the high frequency content of ν_p and it is the quantity of interest in (depth) migration/seismic imaging.

The validity of the mathematics in the following section requires v to satisfy the *geometric optics assumption*: any two points in the medium can be connected by a unique ray, which is a characteristic curve of the eikonal equation; see (6) and Remark 2.1a) below. In this framework, multiple reflections and caustics are precluded.

We further suppose that source and receiver positions are grouped into pairs which can be smoothly parametrized by a variable $s \in S \subset \mathbb{R}$, that is, $\mathbf{x}_r = \mathbf{x}_r(s)$ and $\mathbf{x}_s = \mathbf{x}_s(s)$ are smooth functions of s . The set S is the aperture of the seismic experiment, i.e., the range of source/receiver positions. As we will see in Section 4, the extent of S also determines the migration aperture, that is, the set of reflector dips that can be reconstructed. We assume that S is open and bounded.

Following, e.g., [34] and [16], we can derive that n approximately solves an integral equation, that is,

$$(3) \quad Fn(s, t) \approx \frac{1}{\pi} \int_0^t (\tilde{u} - u)(r; \mathbf{x}_r(s), \mathbf{x}_s(s)) dr.$$

The above right-hand side is the preprocessed seismogram where $u(r; \mathbf{x}_r(s), \mathbf{x}_s(s))$ is the measured reflected wave field at time r and the source/receiver pair given by s . The second wave field $\tilde{u}$ is the reference solution, which is the solution of (1) with ν_p replaced by the macro velocity v . In practical applications, the calculation of $\tilde{u}$ can be omitted when the seismic signal is muted accordingly.

The operator F is an integral operator that acts on n as follows:

$$(4) \quad Fn(s, t) := \int_{\mathcal{R}(s, t)} \frac{A(s, \mathbf{x}) n(\mathbf{x})}{|\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|} d\mathbf{s}(\mathbf{x})$$

where $|\cdot|$ denotes the Euclidean norm. Further,

$$(5) \quad A(s, \mathbf{x}) = \frac{a(\mathbf{x}, \mathbf{x}_s(s)) a(\mathbf{x}, \mathbf{x}_r(s))}{v^2(\mathbf{x})} \quad \text{and} \quad \varphi(s, \mathbf{x}) = \tau(\mathbf{x}, \mathbf{x}_s(s)) + \tau(\mathbf{x}, \mathbf{x}_r(s))$$

where the travel time τ and the amplitude a are solutions of the following partial differential equations,

$$(6) \quad |\nabla_{\mathbf{x}} \tau(\cdot, \mathbf{y})| = \frac{1}{v(\cdot)}, \quad \tau(\mathbf{y}, \mathbf{y}) = 0,$$

and

$$(7) \quad \operatorname{div}_{\mathbf{x}}(a^2(\cdot, \mathbf{y}) \nabla_{\mathbf{x}} \tau(\cdot, \mathbf{y})) = 0, \quad \lim_{\mathbf{x} \rightarrow \mathbf{y}} a(\mathbf{x}, \mathbf{y}) |\mathbf{x} - \mathbf{y}|^{1/2} = \frac{1}{2\pi} \sqrt{\frac{v(\mathbf{y})}{2}}.^2$$

The operator F is a generalized Radon transform since the integration in (4) is carried out along (reflection) isochrones

$$(8) \quad \mathcal{R}(s, t) = \{\mathbf{x} \in X : t = \varphi(s, \mathbf{x})\}$$

which connects points $\mathbf{x}$ whose travel times to source $\mathbf{x}_s(s)$ and receiver $\mathbf{x}_r(s)$ add up to t .

To have Fn be well-defined, we must require two further assumptions. First, we must concretize X as

$$X = \operatorname{int}\{\mathbf{x} \in \mathbb{R}_+^2 : \nabla_{\mathbf{x}} \varphi(s, \mathbf{x}) \neq \mathbf{0} \text{ for all } s \in S\}.$$

So, we exclude any point that lies on a ray connecting a source directly to its receiver, and this ray has a horizontal tangent at that point. Second, n needs to have compact support $\operatorname{supp} n \subset X$, where $\operatorname{supp} n$ is the smallest compact set such that $n = 0$ on $X \setminus \operatorname{supp} n$.

Remark 2.1. a) *Seismic or optical rays are the (bi-)characteristic curves $\mathbf{r}$ of the eikonal equation (6); see, e.g., [3, Appendix E.2.3]. The ray system, where the running parameter $t \geq 0$ agrees with the travel time from the source point, is given by*

$$\frac{d\mathbf{r}}{dt} = v(\mathbf{r}) \frac{\mathbf{p}}{|\mathbf{p}|}, \quad \mathbf{r}(0) = \mathbf{x}_s; \quad \frac{d\mathbf{p}}{dt} = -\frac{\nabla v(\mathbf{r})}{v(\mathbf{r})}, \quad \mathbf{p}(0) = \frac{\boldsymbol{\xi}}{v(\mathbf{r}(0))},$$

with a unit vector $\boldsymbol{\xi} = (\xi_1, \xi_2)^\top \in S^1$. Note that $\mathbf{p}(t) = \nabla \tau(\mathbf{r}(t), \mathbf{x}_s)$.

b) *If the macro velocity model v is linearly increasing with depth, then all (but one) of the rays starting from a fixed surface point resurface. In the common offset measurement configuration with $|\mathbf{x}_s(s) - \mathbf{x}_r(s)| = \operatorname{const}$ for all $s \in S$, rays connecting source and receiver only reach a certain maximal depth $x_{\min} > 0$ where*

$$\nabla_{\mathbf{x}} \tau((x_1, x_{\min})^\top, \mathbf{x}_s(s)) = -\nabla_{\mathbf{x}} \tau((x_1, x_{\min})^\top, \mathbf{x}_r(s)) \quad \text{for all } s \in S \text{ and all } x_1 \in \mathbb{R},$$

see [22, Remark 3.2]. In other words, $X = \{\mathbf{x} \in \mathbb{R}_+^2 : x_2 \neq x_{\min}\}$.

c) *In three spatial dimensions, F is defined analogously to its two-dimensional counterpart: wave propagation is modeled by an analogue of (1) in $\mathbb{R}^3$ and the corresponding approximate equation (3) is again obtained from the ansatz (2). Sources and receivers are situated at $\partial \mathbb{R}_+^3 = \mathbb{R}^2 \times \{0\}$ with positions parametrized by $\mathbf{s} = (s_1, s_2)$ in an open subset (aperture) $S \subset \mathbb{R}^2$. Consequently, $x_3 > 0$ represents depth. The three-dimensional version of F is formulated using the three-dimensional versions of $A = A(\mathbf{s}, \mathbf{x})$ and $\varphi = \varphi(\mathbf{s}, \mathbf{x})$. However, the right-hand side of (3) needs to be modified to*

$$\int_0^t (t-r) (\tilde{u} - u)(r; \mathbf{x}_r(s), \mathbf{x}_s(s)) dr,^3$$

see [17]. The integral in (4) has now transformed into a two-dimensional surface integral over $\mathcal{R}(\mathbf{s}, t) \subset \mathbb{R}_+^3$.

²See Appendix A for a physical justification of the specified asymptotics of a .

³There is indeed no factor in front of the integral essentially due to our normalization of a , see Remark A.1.

In a much simplified setting (among other simplifications, the macro velocity v is a constant), the authors of [5, eqns. (6) and (8)] discovered a three-dimensional variant of (3).

The geophysical term "demigration" refers to the process of numerically evaluating $F_n(s, t)$ for a known n at certain values for s and t . For instance, let $n = \delta_{\mathbf{y}}$ be the Dirac distribution at $\mathbf{y} \in X$. Then,

$$F\delta_{\mathbf{y}}(s, t) = C(s, \mathbf{y}) \delta_{\varphi(s, \mathbf{y})}(t)$$

is the seismogram of a point reflector at $\mathbf{y}$ (C is a function which is not explicitly given in general). The curve in the (s, t) -plane where $F\delta_{\mathbf{y}}(s, t)$ does not vanish is the graph of the function $s \mapsto \varphi(s, \mathbf{y})$, known as the Huygens surface or diffraction travel time with respect to $\mathbf{y}$; see [32, Chap. 2].

The inverse of demigration is true-amplitude migration, which entails recovering n from the measurements:

Given a data set $\{g(s, t) : s \in S, t \in (0, T]\}$ (for instance, the right hand side of (3)), find an n such that

$$Fn(s, t) = g(s, t).$$

All types of Kirchhoff migration formulas known to us, acting on seismogram data $g = g(s, t)$, can be written in our context as

$$KF_W^\dagger(\text{Id} \otimes L)g,$$

see Example 2.2 below for specific examples. We now decompose this formula into its elementary components:

- (a) The operation $\text{Id} \otimes L$ is applied to g where the identity Id acts on the aperture variable s and L on the time variable t . Typically, L is a kind of derivative; for example, $L = \partial_t$ where $(\text{Id} \otimes L)g = \partial_t g$. In any case, the output function $(\text{Id} \otimes L)g$ is seismogram-like depending on (s, t) .
- (b) The operator $F_W^\dagger$ denotes the (generalized) backprojection which transforms a seismogram-like function $h = h(s, t)$ into a function defined over X , according to

$$F_W^\dagger h(\mathbf{x}) = \int_S W(s, \mathbf{x}) h(s, \varphi(s, \mathbf{x})) \, ds$$

where $W : S \times X \rightarrow \mathbb{R}$ is a smooth weight function (or kernel). Note that h is averaged along the Huygens surface with respect to $\mathbf{x} \in X$.

- (c) Finally, K is a kind of partial differential operator with respect to x_1 and x_2 , for instance, $K = -\Delta_{\mathbf{x}}$ (Laplace operator).

Example 2.2. *We list some concrete migration formulas from the literature, expressed in our notation:*

- (i) In [34, Sec. 8], $L = \text{Id}$, $K = \text{Id}$, $W = A$, so that

$$KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{x}) = F_W^\dagger g = \int_S A(s, \mathbf{x}) g(s, \varphi(s, \mathbf{x})) \, ds.$$

Here, $F_W^\dagger$ coincides with F^* , the formal L^2 -adjoint of F .

- (ii) In [1, formula (4.13)], L is a convolution in time with kernel $\lambda(t) = \frac{1}{2\pi} \int_0^\infty \omega e^{it\omega} d\omega$, $K = \frac{1}{2\pi} \text{Id}$, $W = |B|/A$ where

$$B(s, \mathbf{x}) = \det \begin{pmatrix} \nabla_{\mathbf{x}} \varphi(s, \mathbf{x}) \\ \partial_s \nabla_{\mathbf{x}} \varphi(s, \mathbf{x}) \end{pmatrix}$$

is the Beylkin determinant, so that

$$KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{x}) = \frac{1}{2\pi} \int_S \frac{|B(s, \mathbf{x})|}{A(s, \mathbf{x})} (\lambda \star_t g)(s, \varphi(s, \mathbf{x})) \, ds.$$

- (iii) In [3, formula (5.1.40)], $L = H_t \partial_t$ (H_t is the Hilbert transform with respect to t , see Example 3.2 c) below), $K = \frac{1}{2\pi} \text{Id}$, $W = |B|/A$, so that

$$KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{x}) = \frac{1}{2\pi} \int_S \frac{|B(s, \mathbf{x})|}{A(s, \mathbf{x})} (H_t \partial_t g)(s, \varphi(s, \mathbf{x})) \, ds.$$

- (iv) In [3, formula (5.1.41)], $L = H_t \partial_t^2$, $K = \frac{1}{2\pi} \text{Id}$, $W = |B|/(v^2 A |\nabla_{\mathbf{x}} \varphi|)$, so that

$$(9) \quad KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{x}) = \frac{1}{2\pi} \int_S \frac{|B(s, \mathbf{x})|}{v^2(\mathbf{x}) A(s, \mathbf{x}) |\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|} (H_t \partial_t^2 g)(s, \varphi(s, \mathbf{x})) \, ds.$$

- (v) In [13, Sec. 3.2], $L = \text{Id}$, $K = \frac{1}{2\pi} (-\Delta_{\mathbf{x}})^{i/2}$, $i \in \{1, 2\}$, $W = |B|/(A |\nabla_{\mathbf{x}} \varphi|)$, so that

$$KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{x}) = \frac{1}{2\pi} (-\Delta_{\mathbf{x}})^{i/2} \int_S \frac{|B(s, \mathbf{x})|}{A(s, \mathbf{x}) |\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|} g(s, \varphi(s, \mathbf{x})) \, ds.$$

The meaning of $(-\Delta_{\mathbf{x}})^{1/2}$ will be explained below in Example 3.2 b).

- (vi) The authors of [32, Chap. 7, formula (39)] developed a true-amplitude migration formula in three-dimensional space based on ray theory that accounts for caustics. In view of [32, Chap. 7, eqn. (55)], their formula corresponds with (9) in our two-dimensional, caustic-free setting. However, they also address setting $A = 1/v^2$ in case computing the correct amplitude a is too expensive. Then, (9) reduces to

$$KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{x}) = \frac{1}{2\pi} \int_S \frac{|B(s, \mathbf{x})|}{|\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|} (H_t \partial_t^2 g)(s, \varphi(s, \mathbf{x})) \, ds.$$

As a consequence of the provided examples, we conclude that, in order to understand the effect of the different formulas, it is sufficient to study the impact of the operator $KF_W^\dagger(\text{Id} \otimes L)$ on the best possible data g , that is, consistent data $g = \psi F n$. Here, the smooth tapering function $\psi = \psi(s)$ has compact support in S and mitigates boundary effects from limited aperture. A sensible choice of the tapering function ψ is constant, say 1, over a large part of S .

Accordingly, we define the imaging operator

$$(10) \quad \mathcal{L} = \mathcal{L}(K, L, W, F, \psi) := KF_W^\dagger(\text{Id} \otimes L) \psi F,$$

whose mapping properties have been explored in [13] as a pseudodifferential operator, a mathematical concept which we motivate and introduce in the next section.

3. MATHEMATICAL FOUNDATIONS: MICROLOCAL ANALYSIS

This section provides the mathematical notation to understand the geophysical implications of the main result of [13] applied to $\mathcal{L}$, which we present in the next section (Theorem 4.2).

The concepts introduced here belong to the mathematical area of microlocal analysis; see, e.g., [21], which comprises techniques of Fourier/harmonic analysis for investigating partial differential equations. We emphasize to the mathematically experienced reader that we compromise on mathematical precision in favor of a broader perspective and easier accessibility.

An essential tool is the Fourier transform $\widehat{f} = \mathcal{F}f$ of an integrable function $f: \mathbb{R}^d \rightarrow \mathbb{C}$ defined by

$$(11) \quad \widehat{f}(\xi) = \mathcal{F}f(\xi) = \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i\xi \cdot \mathbf{x}} d\mathbf{x}, \quad \xi \in \mathbb{R}^d.$$

If $\widehat{f}$ is integrable as well, we have

$$f(\mathbf{x}) = \mathcal{F}^{-1}\widehat{f}(\mathbf{x}) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{i\xi \cdot \mathbf{x}} d\xi.$$

Both, $\mathcal{F}f$ and $\mathcal{F}^{-1}f$ can be meaningfully defined for f in $L^2(\mathbb{R}^d) = \{v: \mathbb{R}^d \rightarrow \mathbb{C} \text{ measurable} : \int_{\mathbb{R}^d} |v(\mathbf{x})|^2 d\mathbf{x} < \infty\}$, the Hilbert space of square integrable functions. In this case, $\mathcal{F}f$ and $\mathcal{F}^{-1}f$ are in $L^2(\mathbb{R}^d)$.

3.1. Pseudodifferential operators and Sobolev spaces. We now introduce fundamental mathematical concepts that will clarify the performance of each of the migration formulas in Example 2.2. Each of the operators K and L appearing in that example is a pseudodifferential operator. We will use their properties as pseudodifferential operators to understand what object features (or singularities) will be visible in the respective reconstructions. Another mathematical object we will consider is the concept of Sobolev spaces, which provide a rigorous measure of the smoothness of functions, and allow us to say when one reconstruction is smoother than another.

We motivate the definition of pseudodifferential operators by showing they are natural generalizations of differential operators. For the open set $X \subset \mathbb{R}^d$, let $\mathcal{C}^\infty(X)$ be the space of all infinitely differentiable functions on X . Let $P: \mathcal{C}^\infty(X) \rightarrow \mathcal{C}^\infty(X)$ be a linear differential operator of order $m \in \mathbb{N}$,

$$Pu(\mathbf{x}) = \sum_{|\alpha| \leq m} f_\alpha(\mathbf{x}) \partial^\alpha u(\mathbf{x}), \quad f_\alpha \in \mathcal{C}^\infty(X),$$

where the sum is taken over all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = \sum_{i=1}^d \alpha_i \leq m$ and where ∂^α denotes the partial derivative $\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}$.

If u is in $\mathcal{C}_0^\infty(X)$, the subspace of compactly supported functions in $\mathcal{C}^\infty(X)$, then $\widehat{\partial^\alpha u}(\xi) = i^{|\alpha|} \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d} \widehat{u}(\xi) := (i\xi)^\alpha \widehat{u}(\xi)$ and

$$Pu(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\mathbf{x} \cdot \xi} p(\mathbf{x}, \xi) \widehat{u}(\xi) d\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_X e^{i(\mathbf{x}-\mathbf{y}) \cdot \xi} p(\mathbf{x}, \xi) u(\mathbf{y}) d\mathbf{y} d\xi$$

where

$$(12) \quad p(\mathbf{x}, \xi) = \sum_{|\alpha| \leq m} f_\alpha(\mathbf{x}) (i\xi)^\alpha \in \mathcal{C}^\infty(X \times \mathbb{R}^d)$$

is the symbol of P . This representation of P suggests that p can be replaced by more general functions.

Definition 3.1. We call a function $p \in \mathcal{C}^\infty(X \times \mathbb{R}^d \setminus \{0\})$ a **symbol** of order $m \in \mathbb{R}$ if for any $\alpha, \beta \in \mathbb{N}_0^d$ and any compact $K \subset X$ there is a positive constant $C = C_{\alpha, \beta, K}$ such that

$$\sup_{\mathbf{x} \in K} |\partial_{\mathbf{x}}^\alpha \partial_\xi^\beta p(\mathbf{x}, \xi)| \leq C (1 + |\xi|)^{m-|\beta|} \quad \text{for all } \xi \in \mathbb{R}^d \text{ with } |\xi| \geq 1$$

and $\partial_{\mathbf{x}}^\alpha p(\mathbf{x}, \xi)$ is integrable on $K \times (\Omega \setminus \{0\})$ where $\Omega \subset \mathbb{R}^d$ is the unit ball at the origin.

The set of all symbols of order m on X is denoted by $S^m(X)$.

For a $p \in S^m(X)$ we define the operator $\Psi_p: \mathcal{C}_0^\infty(X) \rightarrow \mathcal{C}^\infty(X)$ by

$$\Psi_p u(\mathbf{x}) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_X e^{i(\mathbf{x}-\mathbf{y}) \cdot \xi} p(\mathbf{x}, \xi) u(\mathbf{y}) d\mathbf{y} d\xi$$

and call it a **pseudodifferential operator** of order m . If Ψ_p maps $\mathcal{C}_0^\infty(X)$ into $\mathcal{C}_0^\infty(X)$, we call Ψ_p **properly supported**.

A symbol $\sigma \in S^m(X)$, which is positive homogeneous of order m in its second argument⁴, is a **top order** or **principal symbol** of Ψ_p if there is an $\epsilon > 0$ such that $p - \sigma \in S^{m-\epsilon}(X)$ but $p - \sigma \notin S^m(X)$.

The principal symbol is a key tool for understanding the microlocal structure of a pseudodifferential operator. It encodes the operator's leading-order behavior, especially in the high-frequency or asymptotic regime. Through the principal symbol, one can analyze how operators map high-frequency features, i.e., singularities (see Section 3.2 below). In short, the principal symbol provides fundamental insight into how the operator acts both locally and in phase space.

Example 3.2. a) A representative of the top-order symbol of the differential operator P from (12) is

$$\sigma(x, \xi) = \sum_{|\alpha|=m} f_\alpha(x) (i\xi)^\alpha.$$

b) Given the symbol $|\xi|^2 = \sum_{i=1}^d \xi_i^2 \in S^2(\mathbb{R}^d)$ of the Laplacian $-\Delta = -\sum_{i=1}^d \partial^2 e_i$, we define $(-\Delta)^{1/2}$ as the pseudodifferential operator of order 1 with symbol $|\xi| \in S^1(\mathbb{R}^d)$.

Similarly, $(-\Delta)^{-1/2}$ has symbol $|\xi|^{-1} \in S^{-1}(\mathbb{R}^d)$, $d \geq 2$. Its order is -1 . For $d = 2$ and $u \in \mathcal{C}_0^\infty(\mathbb{R}^2)$ one can show that

$$(-\Delta)^{-1/2} u(\mathbf{x}) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{u(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|} d\mathbf{y}.$$

c) For $u \in \mathcal{C}_0^\infty(\mathbb{R})$, the Hilbert transform is given by the Cauchy principal value integral

$$Hu(x) = \frac{1}{\pi} \oint_{\mathbb{R}} \frac{u(y)}{x - y} dy,$$

which is a pseudodifferential operator of order 0 with symbol $-i \operatorname{sgn}(\xi)$.

To explain the meaning of the order of a pseudodifferential operator, we now introduce Sobolev spaces, which quantify the smoothness of functions.

Definition 3.3. The **Sobolev space** $H^r(\mathbb{R}^d)$, $r \geq 0$, is the space of all $u \in L^2(\mathbb{R}^d)$ such that

$$\|u\|_r^2 := \int_{\mathbb{R}^d} |\widehat{u}(\xi)|^2 (1 + |\xi|^2)^r d\xi < \infty.$$

If K is a compact subset of $\mathbb{R}^d$, then we define $H^r(K)$ to be the subspace of $H^r(\mathbb{R}^d)$ of all functions supported in K .

For $X \subset \mathbb{R}^d$, we define $H_{\text{loc}}^r(X)$ to be the set of all functions v on X such that $\phi v \in H^r(\mathbb{R}^d)$ for every $\phi \in \mathcal{C}_0^\infty(X)$.

⁴We have $p(\mathbf{x}, \lambda\xi) = \lambda^m p(\mathbf{x}, \xi)$ for $\lambda > 0$.

Note that $H^r(\mathbb{R}^d)$ is a Hilbert space with norm $\|\cdot\|_r$ and we have $H^0(\mathbb{R}^d) = L^2(\mathbb{R}^d)$. Further, if $u \in H^r(\mathbb{R}^d)$ and $r \in \mathbb{N}$, then any partial derivative of u of order less than or equal to r is in $L^2(\mathbb{R}^d)$. Sobolev spaces can be defined even for negative indices, and they are subspaces of $\mathcal{S}'$, the space of tempered distributions; see Definition B.1 in the appendix. Tempered distributions can be Fourier transformed; see Definition B.2. So, $H^{-r}(\mathbb{R}^d)$, $r > 0$, is the space of all $u \in \mathcal{S}'$ that have locally integrable Fourier transform $\widehat{u}$ such that $\|u\|_{-r} < \infty$. For instance, the Dirac distribution $\delta_{\mathbf{y}}$ has Fourier transform $e^{-i\mathbf{y} \cdot \boldsymbol{\xi}}$, see (36) below, and is therefore in $H^{-r}(\mathbb{R}^d)$ for any $r > d/2$.

Theorem 3.4. *Let $K \subset X$ be compact. If $p \in S^m(X)$ then $\Psi_p: H^r(K) \rightarrow H_{\text{loc}}^{r-m}(X)$ continuously, that is, for any $\phi \in \mathcal{C}_0^\infty(X)$ there is a constant $C = C_{r,m,\phi}$ such that*

$$\|\phi \Psi_p u\|_{r-m} \leq C \|u\|_r \quad \text{for all } u \in H^r(K).$$

Consequently, a pseudodifferential operator of positive or negative order $\pm m$ ($m > 0$) decreases or increases the Sobolev smoothness of a function by m , respectively. For example, let $u \in L^2(\mathbb{R}^d)$ then $(-\Delta)^{-1/2}u \in H_{\text{loc}}^1(\mathbb{R}^d)$.

3.2. Singularities: wave front sets. The wave front set of a distribution contains its singularities, that is, the locations and the directions where the distribution fails to be smooth. It is essential in microlocal analysis, allowing us to understand how singularities propagate and how pseudodifferential operators, and consequently, migration formulas affect them. We will use this important mathematical concept to analyze which singularities of the ground truth are visible in its reconstruction and how strong they are.

Definition 3.5. i) *We say that $u \in \mathcal{S}'$ is (microlocally) H^s at $(\mathbf{y}, \eta) \in T^* := \mathbb{R}^d \times (\mathbb{R}^d \setminus \{0\})$ if the following holds: for some neighborhood U of $\mathbf{y}$ and some conic neighborhood V of η we have that*

$$\int_V |\widehat{\varphi u}(\xi)|^2 (1 + |\xi|^2)^s d\xi < \infty$$

for one $\varphi \in \mathcal{C}_0^\infty(U)$ with $\varphi(\mathbf{y}) \neq 0$.

ii) *The H^s -wave front set $\text{WF}^s(u)$ of $u \in \mathcal{S}'$ is the complement in T^* of the set of all points $(\mathbf{y}, \eta) \in T^*$ where u is H^s .*

iii) *The pseudo-differential operator Ψ_p with $p \in S^m(X)$ is (microlocally) **elliptic** at $(\mathbf{y}, \eta) \in T^*$, $\mathbf{y} \in X$, if, for some neighborhood $U \subset X$ of $\mathbf{y}$, some conic neighborhood V of η , there exist $C > 0$ and $M > 0$ such that*

$$|p(\mathbf{x}, \xi)| \geq C |\xi|^m \quad \text{for } |\xi| \geq M, \xi \in V, \text{ and for all } \mathbf{x} \in U.$$

(Note that ellipticity of Ψ_p is a property of its principal symbol.)

Note that $\text{WF}^t(u) \subset \text{WF}^s(u)$ for $t < s$ because the weights satisfy $(1 + |\xi|^2)^t \leq (1 + |\xi|^2)^s$. The set of all singularities of u is the *wave front set*

$$\text{WF}(u) := \text{Cl}\left(\bigcup_{s \in \mathbb{R}} \text{WF}^s(u)\right).^5$$

Roughly speaking, $(\mathbf{x}, \xi) \in \text{WF}(u)$ if u fails to be a $\mathcal{C}^\infty$ -function at $\mathbf{x}$ in the direction ξ . Ψ DOs do not create singularities: $\text{WF}(Pu) \subset \text{WF}(u)$ for any Ψ DO P .

Example 3.6. *Let $\mathbf{1}_\Omega$ be the characteristic function of a domain $\Omega \subset \mathbb{R}^2$ with a $\mathcal{C}^\infty$ -boundary. Then, for any $\epsilon \geq 0$,*

$$\text{WF}^{1/2+\epsilon}(\mathbf{1}_\Omega) = \{(\mathbf{x}, \xi) : \mathbf{x} \in \text{bd}(\Omega), \xi \in \mathbb{R}^d \setminus \{0\}, \xi \perp \text{bd}(\Omega) \text{ at } \mathbf{x}\}$$

⁵ $\text{Cl}(A)$ denotes the closure of the set A .

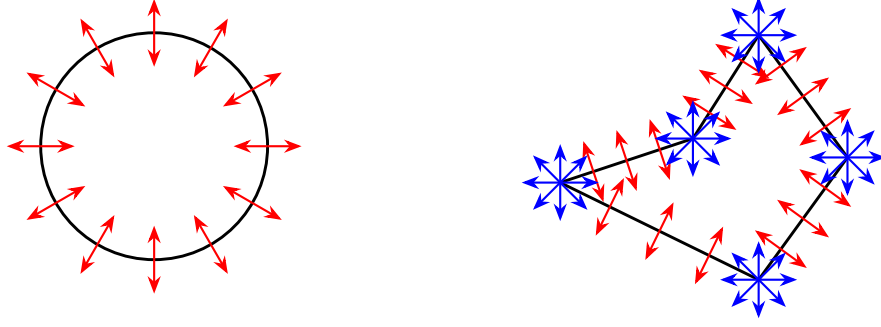


FIGURE 1. Illustrations of the wave front sets of the indicator functions of a circular disk (left) and a non-convex pentagon (right).

and, for any $\gamma > 0$,

$$\text{WF}^{1/2-\gamma}(\mathbf{1}_\Omega) = \emptyset.$$

Hence, $\text{WF}(\mathbf{1}_\Omega) = \text{WF}^{1/2+\epsilon}(\mathbf{1}_\Omega)$, see [15, Examp. 2.24].

A similar result holds for the characteristic function $\mathbf{1}_P$ of a polygon $P \subset \mathbb{R}^2$ except for corner points $\mathbf{c} \in \text{bd}(P)$ where $(\mathbf{c}, \xi) \in \text{WF}^{1/2}(\mathbf{1}_P)$ for any $\xi \in \mathbb{R}^2 \setminus \{0\}$, see [11, Example 3.10].

Both examples are illustrated in Figure 1.

The following theorem gives a detailed description of how a Ψ DO maps wave fronts. Depending on its order m , an elliptic Ψ DO emphasizes ($m > 0$) or de-emphasizes ($m < 0$) singularities. In any case, it conserves the singularities in directions where it is elliptic.

Theorem 3.7. *Let Ψ_p be a pseudodifferential operator with $p \in S^m(X)$ that is elliptic at $(\mathbf{y}, \eta)$. Then, for any u with compact support (Definition B.3),*

$$(\mathbf{y}, \eta) \in \text{WF}^s(u) \iff (\mathbf{y}, \eta) \in \text{WF}^{s-m}(\Psi_p u).$$

If Ψ_p is properly supported, then the equivalence holds for each $u \in \mathcal{S}'$.

Proof. The theorem is included implicitly in [26]. An explicit proof is provided in [15, Thm. 2.23] based on [27, Thm. 3.1]. $\square$

Example 3.8. Recall the Ψ DO $(-\Delta)^{-1/2}$ of order -1 from Example 3.2 b). It is easy to see that it is microlocally elliptic at any point in T^* . So, for any $s \in \mathbb{R}$ we have that

$$\text{WF}^{s+1}((-\Delta)^{-1/2}u) = \text{WF}^s(u)$$

and $\text{WF}((-\Delta)^{-1/2}u) = \text{WF}(u)$.

As can be seen in Figure 2, the wave fronts are kept and only slightly smoothed. Please observe that the corners of both rectangles are clearly visible as proper angles, since $(\mathbf{c}, \xi) \in \text{WF}^{3/2}((-\Delta)^{-1/2}u)$ for any corner point $\mathbf{c}$ and any direction ξ ; compare Figure 1 (right).

4. THE GENERAL IMAGING OPERATOR AND ITS PRINCIPAL SYMBOL

One important technical requirement must be mentioned before we can formulate the principal symbol of the imaging operator $\mathcal{L}$ from (10). For $\mathcal{L}$, which is a product of different operators, to be a well-defined Ψ DO, the Bolker condition [18, 19] must hold on X , which is an implicit restriction on the macro velocity v restricted to X in combination with the source/receiver configuration (measurement geometry). The origin and exact definition of this condition are beyond the scope of this exposition, as it would significantly increase the mathematical overhead

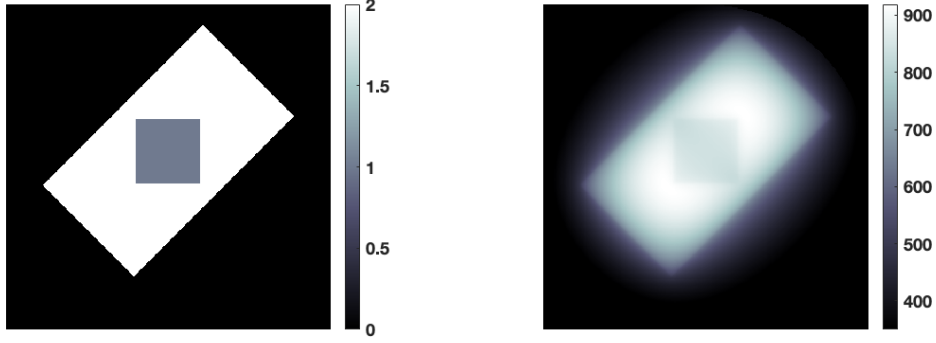


FIGURE 2. The function u as a sum of two indicator functions of rectangles (left) and a numerical approximation of $(-\Delta)^{-1/2}u$ (right). The equality $\text{WF}^{1/2}(u) = \text{WF}^{3/2}((-\Delta)^{-1/2}u)$ is clearly visible.

without yielding deeper insight (but see Remark 4.1a) below). In fact, there is no simple way to verify this condition. However, a few examples are known, and it is stable insofar as small, compactly supported perturbations of the macro velocity do not affect it; see [22, Sec. 5.3].

The Bolker condition holds for

- a constant macro velocity on any X with common offset or zero offset configuration [8], and
- an affine-linear macro velocity with common offset or zero offset configuration (provided X keeps a certain distance from the surface) [22].

On the other hand, the Bolker condition is violated for a constant macro velocity with the common midpoint configuration [8].

Remark 4.1. a) *For the inquiring reader, we formulate the Bolker condition in our context. Define the mapping*

$$\Pi: \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \times X \rightarrow \mathbb{R}^4, \quad (s, \omega, \mathbf{x}) \mapsto (s, \varphi(s, \mathbf{x}), -\omega \partial_s \varphi(s, \mathbf{x}), \omega), \quad \text{with } \varphi \text{ from (5).}$$

The Bolker condition holds in X if both the mapping and its Jacobian are injective; see [22, Sec. 3].

b) *If the Bolker condition is violated, the migration formulas in Example 2.2 become unreliable because they may introduce singularities that are not present in the ground truth. Though these added artifacts can be characterized to a certain extent, see [28].*

Theorem 4.2. *Let X and S be bounded open sets with compact closures and assume F satisfies the Bolker condition on X . Let K be a properly supported ΨDO on $\mathbb{R}^2$ of order κ with principal symbol $k = k(\mathbf{x}, \xi)$, which is bounded at $\xi = 0$, and let L be a ΨDO on $\mathbb{R}$ of order λ with principal symbol $\ell = \ell(\omega)$. Let $\psi \in \mathcal{C}_0^\infty(S)$. Then, the operator*

$$\mathcal{L} = K F_W^\dagger (\text{Id} \otimes L) \psi F$$

is a well-defined ΨDO on X of order $\kappa + \lambda - 1$. Its principal symbol is

$$\sigma(\mathcal{L})(\mathbf{x}, \xi) = 2\pi \frac{k(\mathbf{x}, \xi)}{|\xi|} \sum_{(s, \omega) \in \mathcal{I}(\mathbf{x}, \xi)} \frac{\psi(s) W(s, \mathbf{x}) \ell(\omega) A(s, \mathbf{x}) |\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|}{|B(s, \mathbf{x})|}$$

where

$$\mathcal{I}(\mathbf{x}, \xi) = \{(s, \omega) \in \text{Cl}(S) \times (\mathbb{R} \setminus \{0\}) : \xi = \omega \nabla_{\mathbf{x}} \varphi(s, \mathbf{x})\}$$

is either finite or empty. In the latter case, $\sigma(\mathcal{L})(\mathbf{x}, \xi) = 0$.

Proof. See Lemma 3.1 and Theorem 3.2 in [13]. Note that the normalizations of $\mathcal{L}$ and the symbols in [13] differ from ours. $\square$

We will use Theorem 4.2 to calculate the symbols of the various migration operators. The set $\mathcal{I}(\mathbf{x}, \xi)$ is simpler than it looks; for constant velocity, it consists of at most one point, and for affine-linear macro velocity, it consists of at most two, as seen from (16) below.

In the following subsections, we discuss and compare some of the migration formulas from Example 2.2 in the light of the above theorem and of Theorem 3.7. For convenience, we define the set, for $\mathbf{x} \in X$,

$$(13) \quad E(\mathbf{x}) = \left\{ \xi \in \mathbb{R}^2 \setminus \{0\} : \sum_{(s, \omega) \in \mathcal{I}(\mathbf{x}, \xi)} \psi(s) > 0 \right\}.$$

As we will see below, this set contains the directions in which the imaging operators associated with the migration formulas in Example 2.2 are microlocally elliptic at $\mathbf{x}$.

For future reference we introduce the zero ($\alpha = 0$) and the common offset ($\alpha > 0$) measurement configurations where

$$(14) \quad \mathbf{x}_s(s) = (s - \alpha, 0)^\top \quad \text{and} \quad \mathbf{x}_r(s) = (s + \alpha, 0)^\top,$$

and the macro velocity model

$$(15) \quad v(\mathbf{x}) = a x_2 + b, \quad b > 0, \quad a \geq 0.$$

For these settings, the geometric optics assumption and the Bolker condition hold in

$$X = \{\mathbf{x} \in \mathbb{R}^2 : x_2 > x_{\min}\}, \quad x_{\min} := \begin{cases} \frac{b}{a} \left(\sqrt{1 + \frac{a^2 \alpha^2}{b^2}} - 1 \right) & : a > 0, \\ 0 & : a = 0. \end{cases}$$

Further,

$$(16) \quad \text{zero offset } (\alpha = 0) \implies \sum_{(s, \omega) \in \mathcal{I}(\mathbf{x}, \xi)} \psi(s) = \begin{cases} \psi\left(x_1 - \frac{\xi_1}{\xi_2} x_2\right) & : a = 0, \xi_2 \neq 0, \\ 0 & : a = 0, \xi_2 = 0, \\ \psi(x_1) & : a > 0, \xi_1 = 0, \\ \psi(s_-) + \psi(s_+) & : a > 0, \xi_1 \neq 0, \end{cases}$$

where

$$s_{\pm} = s_{\pm}(\mathbf{x}, \xi) = x_1 + \frac{\xi_2}{\xi_1} \frac{b + a x_2}{a} \pm \sqrt{\frac{x_2}{a} (2b + a x_2) + \left(\frac{\xi_2}{\xi_1} \frac{b + a x_2}{a} \right)^2},$$

see [22, Sec 3.3].

4.1. The normal operator. The imaging operator associated with the Kirchhoff formula from Example 2.2 (i) is the normal operator

$$F_A^\dagger \psi F = F^* \psi F.$$

We start with this example because the normal operator is a fundamental component of the other imaging operators that will be examined below. These results will be readily transferrable to those other cases.

Having $k = \ell = 1$, the principal symbol of the normal operator is

$$(17) \quad \sigma(F^*\psi F)(\mathbf{x}, \xi) = \frac{2\pi}{|\xi|} \sum_{(s, \omega) \in \mathcal{I}(\mathbf{x}, \xi)} \frac{\psi(s) A^2(s, \mathbf{x}) |\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|}{|B(s, \mathbf{x})|}.$$

4.1.1. *Constant macro velocity and zero offset.* For $v(\cdot) = b > 0$ and $\alpha = 0$ we simplify the symbol further. Here, τ and a from (6) and (7) are known explicitly, so that $\varphi(s, \mathbf{x}) = 2\tau(\mathbf{x}, \mathbf{x}_s(s)) = 2|\mathbf{x} - \mathbf{x}_s(s)|/b$ and $a(\mathbf{x}, \mathbf{x}_s(s)) = (8\pi^2|\mathbf{x} - \mathbf{x}_s(s)|/b)^{-1/2}$. Hence,

$$|\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})| = \frac{2}{b}, \quad A^2(s, \mathbf{x}) = \frac{1}{64\pi^4 b^2 |\mathbf{x} - \mathbf{x}_s(s)|^2}, \quad \text{and} \quad B(s, \mathbf{x}) = \frac{4x_2}{b^2 |\mathbf{x} - \mathbf{x}_s(s)|^2},$$

so that

$$\sigma(F^*\psi F)(\mathbf{x}, \xi) \stackrel{(16)}{=} \begin{cases} \frac{1}{64\pi^3 b} \frac{1}{|\xi| x_2} \psi\left(x_1 - \frac{\xi_1}{\xi_2} x_2\right) & : \xi_2 \neq 0, \\ 0 & : \xi_2 = 0. \end{cases}$$

If $\psi(s) > 0$ for $s \in (s_{\min}, s_{\max})$, $s_{\min} < s_{\max}$, then E from (13) is explicitly given by

$$E(\mathbf{x}) = \left\{ \xi \in \mathbb{R}^2 : \xi_2 \neq 0, \frac{x_1 - s_{\max}}{x_2} < \frac{\xi_1}{\xi_2} < \frac{x_1 - s_{\min}}{x_2} \right\},$$

which is a cone that excludes a conical neighborhood of the horizontal axis.

If $\xi \in E(\mathbf{x})$ then $F^*\psi F$ is microlocally elliptic at $(\mathbf{x}, \xi)$ and

$$(18) \quad (\mathbf{x}, \xi) \in \text{WF}^s(u) \iff (\mathbf{x}, \xi) \in \text{WF}^{s+1}(F^*\psi F u).$$

This means that if $\xi \in E(\mathbf{x})$, then u has a singularity at $(\mathbf{x}, \xi)$ exactly when $F^*\psi F u$ has a singularity at $(\mathbf{x}, \xi)$. Furthermore, that singularity is one order smoother in Sobolev scale than the corresponding singularity of u . A similar result holds for the common offset configuration ($\alpha > 0$); see [16, Cor. 4.3].

Note that singularities with horizontal directions ($\xi_2 = 0$) cannot be reconstructed (in fact, $F^*\psi F u(\mathbf{x})$ is $\mathcal{C}^\infty$ in directions $\xi \notin E(\mathbf{x})$ [29, Rem. 3.3]). Moreover, wave front directions near the boundary of $E(\mathbf{x})$ might be reconstructed more weakly than those corresponding to where $\psi = 1$. Further, as the ellipticity of $F^*\psi F$ at $(\mathbf{x}, \xi)$ deteriorates with increasing depth variable x_2 , the wave fronts farther down are reconstructed more weakly than those closer to the surface. However, this aesthetic flaw can be easily corrected: The principal symbol of $M F^*\psi F$, where M is the operator of multiplication by the depth coordinate, $M u(\mathbf{x}) = x_2 u(\mathbf{x})$, does not contain the factor $1/x_2$ anymore. Then, $M F^*\psi F$ behaves essentially like $(-\Delta)^{-1/2}$ at $(\mathbf{x}, \xi)$ where $\xi \in E(\mathbf{x})$, compare Example 3.8.

4.1.2. *Modified normal operators.* Now, we will consider arbitrary macro velocities and measurement configurations that, however, satisfy the Bolker condition. In this case, computing a from (7) might be prohibitively costly. Therefore, we propagate using the weight

$$(19) \quad W_\vartheta(s, \mathbf{x}) = \frac{v^2(\mathbf{x}) |B(s, \mathbf{x})|}{|\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|} \vartheta(s, \mathbf{x}), \quad \vartheta(s, \mathbf{x}) := 8\pi^2 \sqrt{\frac{|\mathbf{x} - \mathbf{x}_s(s)|}{v(\mathbf{x})} \frac{|\mathbf{x} - \mathbf{x}_r(s)|}{v(\mathbf{x})}},$$

all ingredients of which are known, since the travel time τ has to be computed anyway. Then,

$$(20) \quad \sigma(F_{W_\vartheta}^\dagger \psi F)(\mathbf{x}, \xi) = \frac{2\pi}{|\xi|} \sum_{(s, \omega) \in \mathcal{I}(\mathbf{x}, \xi)} \psi(s) \Theta(s, \mathbf{x})$$

where

$$(21) \quad \Theta(s, \mathbf{x}) := a(\mathbf{x}, \mathbf{x}_s(s)) a(\mathbf{x}, \mathbf{x}_r(s)) \vartheta(s, \mathbf{x}) > 0.^6$$

Our definition of the correction factor ϑ in (19) is based on the following rationale: In view of (34) and the subsequent discussion in Appendix A we anticipate that

$$a(\mathbf{x}, \mathbf{y}) \approx \frac{1}{2\pi} \sqrt{\frac{v(\mathbf{x})}{2|\mathbf{x} - \mathbf{y}|}} \quad \text{for } \mathbf{y} \in \{\mathbf{x}_s, \mathbf{x}_r\}.$$

Thus, we expect Θ to attain values around 1, largely independent of s and $\mathbf{x}$ (if $v(\cdot) = b > 0$ then $\Theta = 1$). The bottom of Figure 3 displays a numerical approximation⁷ of the difference $1 - \Theta(s, \cdot)$ for a fixed s . As macro velocity, we used the Marmousi pressure wave distribution [35], a widely used benchmark model in geophysics. Even for this complex velocity model, $\Theta(s, \cdot)$ is close to 1 throughout (as can be seen in the bottom picture of Figure 3).

Remark 4.3. *In accordance with Remark A.1, the appropriate correction factor to be used in three dimensions is*

$$\vartheta(s, \mathbf{x}) = 16\pi^2 |\mathbf{x} - \mathbf{x}_s(s)| |\mathbf{x} - \mathbf{x}_r(s)|.$$

Should a and hence A be available, the weight of choice should be

$$W_A(s, \mathbf{x}) = \frac{|B(s, \mathbf{x})|}{A(s, \mathbf{x}) |\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|}$$

yielding

$$(22) \quad \sigma(F_{W_A}^\dagger \psi F)(\mathbf{x}, \xi) = \frac{2\pi}{|\xi|} \sum_{(s, \omega) \in \mathcal{J}(\mathbf{x}, \xi)} \psi(s).$$

Then, $F_{W_\vartheta}^\dagger \psi F$ and $F_{W_A}^\dagger \psi F$ are microlocally elliptic at $(\mathbf{x}, \xi)$, $\xi \in E(\mathbf{x})$, and (18) holds accordingly for both operators.

Remark 4.4. *Recall that a reasonable choice of the tapering function ψ is constant 1 over a large part of the aperture S . Then, the sum in (22) is constant over a large subset X_o of X , so that the symbols of $F_{W_\vartheta}^\dagger \psi F$ and $F_{W_A}^\dagger \psi F$ are independent of $\mathbf{x} \in X_o$. In contrast, the symbol of $F^* \psi F$, see (17), generically depends heavily on $\mathbf{x} \in X_o$. So, using microlocal analysis, we came up with weights providing reconstructions that are less affected by the location of $\mathbf{x}$. This is most clearly seen in Corollary 4.5 below.*

4.1.3. Linearly increasing macro velocity and zero offset. A fundamental difference exists between constant $a = 0$ and linearly increasing macro velocities $a > 0$; see (15). For $a > 0$ and $\alpha = 0$ we have that

$$E(\mathbf{x}) = E_+(\mathbf{x}) \cup E_-(\mathbf{x}) \cup \{\xi \in \mathbb{R}^2 : \xi_1 = 0, \psi(x_1) > 0\}$$

where $E_\pm(\mathbf{x}) = \{\xi \in \mathbb{R}^2 : \xi_1 \neq 0, \psi(s_\pm) > 0\}$ by (13) and (16). According to [22, Prop. 3.5] the three imaging operators $F^* \psi F$, $F_{W_\vartheta}^\dagger \psi F$, and $F_{W_A}^\dagger \psi F$ are microlocally elliptic at $(\mathbf{x}, \xi)$, $\xi \in E(\mathbf{x})$, and (18) applies to these operators as well.

Observe that, depending on the aperture S and the tapering function ψ , $E(\mathbf{x})$ can contain any direction $\xi \in \mathbb{R}^2 \setminus \{0\}$, so even singularities with a horizontal normal direction can be

⁶Given the assumptions on v in Section 4.1 of [22], especially assumptions (4.4) and (4.5), we can argue that a is positive under the normalization in (7). See the middle image in Figure 3 for a numerical example.

⁷The numerical approximations shown in Figure 3 have been obtained by routines published in [10].

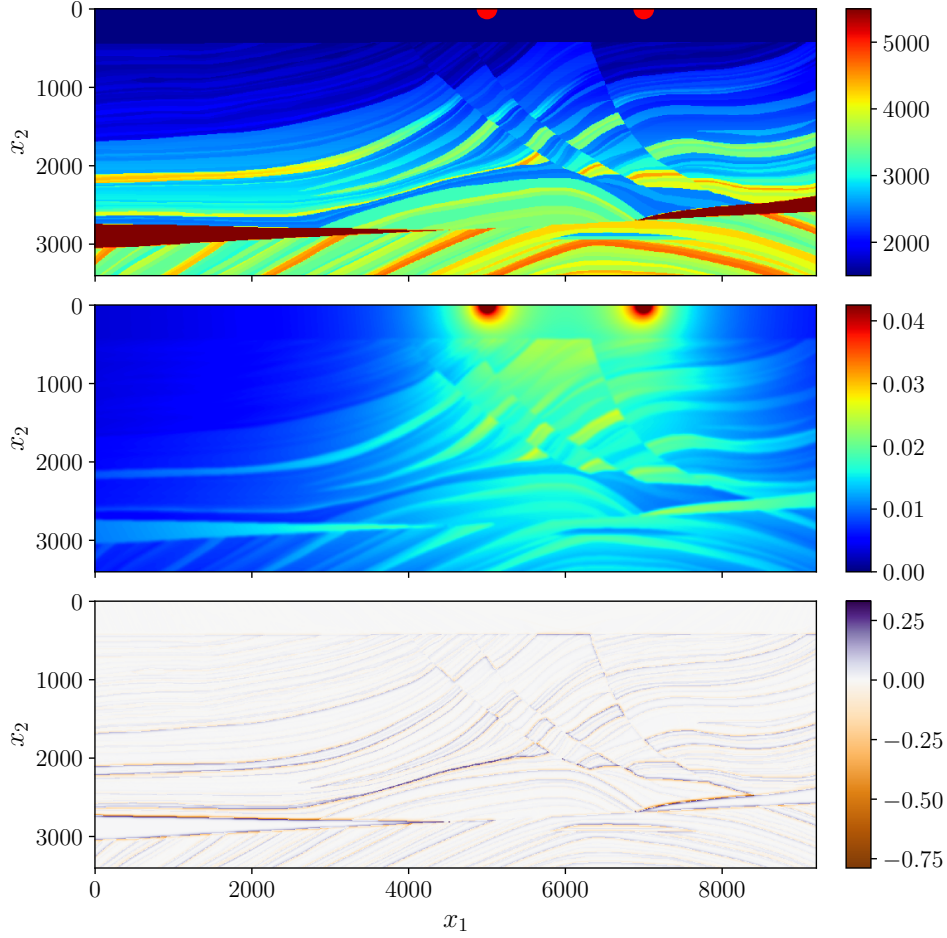


FIGURE 3. Top: Marmousi pressure wave velocity model. The length unit is 1m and the velocity unit is 1m/s. Middle: the corresponding product $a(\cdot, \mathbf{x}_s)a(\cdot, \mathbf{x}_r)$ of the amplitude functions (7), where $\mathbf{x}_s$ and $\mathbf{x}_r$ are 2km apart, and their positions are indicated by red dots in the top image. Bottom: The difference $1 - \Theta(s, \cdot)$ where Θ is from (21) and s is the aperture variable corresponding to $\mathbf{x}_s$ and $\mathbf{x}_r$. As argued in the text below (21), Θ attains values largely around 1.

reconstructed. This is in sharp contrast to the situation of a constant macro velocity where the horizontal direction is not in $E(\mathbf{x})$ regardless of the choice of S and ψ .

4.2. Differentiated normal operators. The operators

$$\Lambda_A^{(i)} := \frac{1}{2\pi} (-\Delta_{\mathbf{x}})^{i/2} F_{W_A}^\dagger \psi F, \quad i \in \{1, 2\},$$

belong to the migration formulas of Example 2.2 (v). They are closely related to the modified normal operators. In view of Theorem 4.2,

$$\sigma(\Lambda_A^{(i)})(\mathbf{x}, \xi) = |\xi|^i \sigma(F_{W_A}^\dagger \psi F)(\mathbf{x}, \xi) = |\xi|^{i-1} \sum_{(s, \omega) \in \mathcal{I}(\mathbf{x}, \xi)} \psi(s).$$

So, the operators, $F_{W_A}^\dagger \psi F$ and $\Lambda_A^{(i)}$, share the same ellipticity properties, while their orders are -1 and $i - 1$, respectively. Furthermore, let $\xi \in E(\mathbf{x})$, then

$$(23) \quad (\mathbf{x}, \xi) \in \text{WF}^s(u) \iff (\mathbf{x}, \xi) \in \text{WF}^{s-(i-1)}(\Lambda_A^{(i)} u).$$

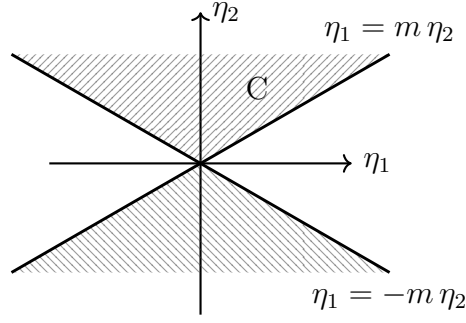


FIGURE 4. The hatched region illustrates an admissible cone C from Corollary 4.5. As m increases, the cone includes more and more directions, but never contains the horizontal η_1 -axis. In general, a large m correlates with a large aperture S . The distributions for which (24) holds must have their singularities in directions in the cone: $\xi \in C$.

The strength of such singularities is either preserved ($i = 1$) or emphasized ($i = 2$).

Now, let $v(\cdot) = b > 0$ and $\alpha = 0$. Then, by (16) we conclude that

$$\sigma(\Lambda_A^{(1)})(\mathbf{x}, \xi) = 1 \quad \text{and} \quad \sigma(\Lambda_A^{(2)})(\mathbf{x}, \xi) = |\xi| \quad \text{whenever} \quad \psi\left(x_1 - \frac{\xi_1}{\xi_2} x_2\right) = 1.$$

Hence, it appears that $\Lambda_A^{(1)}$ is microlocally close to the identity (true-amplitude migration!) and $\Lambda_A^{(2)}$ is close to $(-\Delta_{\mathbf{x}})^{1/2}$. The following corollary mathematically formulates this observation rigorously.

Corollary 4.5. *Let $v(\cdot) = b > 0$ and assume the zero-offset or common offset measurement configuration. To any open, bounded $\mathcal{X} \subset X$ and any cone $C = \{\eta \in \mathbb{R}^2 \setminus \{0\} : |\eta_1| \leq m|\eta_2|\}$, $m > 0$ (see Figure 4), there are an aperture S , a tapering function $\psi \in \mathcal{C}_0^\infty(S)$, and pseudodifferential operators $O^{(i)}$, $i \in \{1, 2\}$, such that*

$$(24) \quad \Lambda_A^{(1)}u = u + O^{(1)}u \quad \text{and} \quad \Lambda_A^{(2)}u = (-\Delta)^{1/2}u + O^{(2)}u,$$

for all compactly supported $u \in \mathcal{S}'$ with $\text{WF}(u) \subset \mathcal{X} \times C$. For those u , $O^{(1)}u$ is one degree smoother in Sobolev scale than u , that is, if $u \in H^s(\text{Cl}(\mathcal{X}))$ then $O^{(1)}u \in H_{\text{loc}}^{s+1}(X)$. Similarly, $O^{(2)}u$ is one degree smoother in Sobolev scale than $(-\Delta)^{1/2}u$.

Proof. For the zero-offset measurement configuration, this result is proven in [13, Cor. 3.7]. However, the proof provided readily carries over to the common offset configuration. In fact, the principal symbols of both operators, for $\alpha \geq 0$, can be expressed as

$$\sigma(\Lambda_A^{(i)})(\mathbf{x}, \xi) = |\xi|^{i-1} \psi\left(x_1 - Q\left(\frac{\xi_1}{\xi_2}, \frac{\alpha}{x_2}\right)x_2\right), \quad i \in \{1, 2\},$$

where

$$Q(q, \lambda) := \frac{1}{2q} \left(q^2 - 1 + \sqrt{(q^2 + 1)^2 + 4\lambda^2 q^2} \right),$$

see [16, Sec. 4.1]. Now, the proof of [13, Cor. 3.7] extends to this representation, even when $\alpha > 0$. $\square$

Should a be unknown then we equip the generalized backprojection with weight W_ϑ (19) and obtain the modified differentiated normal operators

$$(25) \quad \Lambda_\vartheta^{(i)} := \frac{1}{2\pi} (-\Delta_{\mathbf{x}})^{i/2} F_{W_\vartheta}^\dagger \psi F, \quad i \in \{1, 2\},$$

with top order symbols

$$(26) \quad \sigma(\Lambda_\vartheta^{(i)})(\mathbf{x}, \xi) = |\xi|^{i-1} \sum_{(s, \omega) \in \mathcal{J}(\mathbf{x}, \xi)} \psi(s) \Theta(s, \mathbf{x})$$

where Θ is as in (21). Similarly, (23) holds for $\Lambda_\vartheta^{(i)}$ as well. Under the assumptions of Corollary 4.5 we likewise have that

$$(27) \quad \Lambda_\vartheta^{(i)} u = \Psi_{p^{(i)}} u + \tilde{O}^{(i)} u \quad \text{with } p^{(i)}(\mathbf{x}, \xi) = |\xi|^{i-1} \Theta\left(x_1 - \frac{\xi_1}{\xi_2} x_2, \mathbf{x}\right).$$

The symbol $p^{(i)}$ is of order $i - 1$ and $\tilde{O}^{(i)} u$ is one degree smoother in Sobolev scale than $\Psi_{p^{(i)}} u$. As previously argued, Θ remains constant over a large region, implying that $\Psi_{p^{(i)}}$ is approximately a scalar multiple of $(-\Delta_{\mathbf{x}})^{(i-1)/2}$.

Remark 4.6. We highlight a significant difference between the migration formulas with respect to $\Lambda^{(1)}$ and $\Lambda^{(2)}$, where $\Lambda^{(i)}$ represents either $\Lambda_A^{(i)}$ or $\Lambda_\vartheta^{(i)}$: After backprojecting the data we need to apply either $(-\Delta)^{1/2}$ ($i = 1$) or $-\Delta$ ($i = 2$). While the latter operator is a classical differential operator, which is straightforwardly implemented, the former is a pseudodifferential operator that is not properly supported and therefore requires special treatment when implemented; see [13, Rem. 4.1].

4.3. The classical Kirchhoff operators. We refer to the imaging operators belonging to the migration formulas in Examples 2.2 (iii), (iv), and (vi) as classical Kirchhoff operators. These formulas are of filtered backprojection type, since the data undergoes filtering (by the Hilbert transform) prior to backprojection.

First, we consider the variants with a known a . The corresponding operators are

$$I_K^{(1)} := \frac{1}{2\pi} F_{|B|/A}^\dagger (\text{Id} \otimes (H_t \partial_t)) \psi F \quad \text{and} \quad I_K^{(2)} := \frac{1}{2\pi} F_{|B|/(v^2 A |\nabla_{\mathbf{x}} \varphi|)}^\dagger (\text{Id} \otimes (H_t \partial_t^2)) \psi F.$$

They have orders 0 and 1, respectively. Since $\sigma(H_t \partial_t)(t, \omega) = |\omega|$ as well as $\sigma(H_t \partial_t^2)(t, \omega) = i|\omega|\omega$, the principal symbols read

$$\sigma(I_K^{(1)})(\mathbf{x}, \xi) = \sum_{(s, \omega) \in \mathcal{J}(\mathbf{x}, \xi)} \psi(s)$$

and

$$\sigma(I_K^{(2)})(\mathbf{x}, \xi) = \frac{|\xi|}{v(\mathbf{x})^2} \sum_{(s, \omega) \in \mathcal{J}(\mathbf{x}, \xi)} \psi(s) \frac{i \operatorname{sgn}(\omega)}{|\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|^2}.$$

By $\sigma(I_K^{(1)}) = \sigma(\Lambda_A^{(1)})$, Corollary 4.5 (equality on the left in (24)) holds accordingly for $I_K^{(1)}$.

Now, we investigate the principal symbol of $I_K^{(2)}$. Recalling the symbol of the Hilbert transform H_t , $\sigma(H_t)(t, \omega) = -i \operatorname{sgn}(\omega)$ (Example 3.2 c)), we observe that H_t contributes nothing to the migration process because it has order 0 and its symbol appears almost unaltered in $\sigma(I_K^{(2)})$. Consequently, we can eliminate H_t from this migration formula without losing any information. Furthermore, we have v as well as $\nabla_{\mathbf{x}} \varphi$ at hand, which we can use in the weight function

to correct for the corresponding factors. Hence,

$$\tilde{\mathbf{I}}_K^{(2)} = \frac{1}{2\pi} F_{|B||\nabla_{\mathbf{x}}\varphi|/A}^\dagger (\text{Id} \otimes (-\partial_t^2)) \psi F$$

with principal symbol

$$\sigma(\tilde{\mathbf{I}}_K^{(2)})(\mathbf{x}, \xi) = \sigma(\Lambda_A^{(2)})(\mathbf{x}, \xi) = |\xi| \sum_{(s, \omega) \in \mathcal{J}(\mathbf{x}, \xi)} \psi(s)$$

is our recommended substitute for the classical $\mathbf{I}_K^{(2)}$.

In conclusion, we can expect similar migration results from $\Lambda_A^{(1)}$ and $\mathbf{I}_K^{(1)}$, as well as from $\Lambda_A^{(2)}$ and $\tilde{\mathbf{I}}_K^{(2)}$.

Finally, we briefly address the situation where a is not available. Then, our suggested modifications of $\mathbf{I}_K^{(1)}$ and $\tilde{\mathbf{I}}_K^{(2)}$ are

$$(28) \quad \mathbf{I}_\vartheta^{(1)} := \frac{1}{2\pi} F_{|B|v^2\vartheta}^\dagger (\text{Id} \otimes (H_t \partial_t)) \psi F \quad \text{and} \quad \tilde{\mathbf{I}}_\vartheta^{(2)} := \frac{1}{2\pi} F_{|B|v^2|\nabla_{\mathbf{x}}\varphi|\vartheta}^\dagger (\text{Id} \otimes (-\partial_t^2)) \psi F,$$

see (19) for ϑ . They possess the same top order symbols as $\Lambda_\vartheta^{(1)}$ and $\Lambda_\vartheta^{(2)}$, respectively, see (26), and accordingly yield similar results.

4.4. Beylkin's operator. The Beylkin migration formula is presented in Example 2.2 (ii) with imaging operator

$$\mathbf{I}_B := \frac{1}{2\pi} F_{|B|/A}^\dagger (\text{Id} \otimes L) \psi F.$$

Here, L is convolution in time with convolution kernel $\lambda(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \chi(\omega) |\omega| e^{it\omega} d\omega$ where χ is the Heaviside function: $\chi(\omega) = 0, \omega < 0$, and $\chi(\omega) = 1$, otherwise. Thus, $\ell(\omega) = \chi(\omega) |\omega|$ is the symbol of L yielding the order 0 of $\mathbf{I}_B$ and

$$\sigma(\mathbf{I}_B)(\mathbf{x}, \xi) = \sum_{(s, \omega) \in \mathcal{J}(\mathbf{x}, \xi)} \psi(s) \chi(\omega).$$

For an interpretation of the symbol, we consider $v(\cdot) = \mathbf{b}$ with the zero-offset configuration. In this combination, the set $\mathcal{J}(\mathbf{x}, \xi)$, $\mathbf{x} \in X$, contains a single element (s, ω) where s is explicitly given in (16). It can be shown that ω and ξ_2 have the same sign which results in

$$\sigma(\mathbf{I}_B)(\mathbf{x}, \xi) \stackrel{(16)}{=} \begin{cases} \psi\left(x_1 - \frac{\xi_1}{\xi_2} x_2\right) & : \xi_2 > 0, \\ 0 & : \text{otherwise.} \end{cases}$$

So, either $\sigma(\mathbf{I}_B)(\mathbf{x}, \xi)$ or $\sigma(\mathbf{I}_B)(\mathbf{x}, -\xi)$ must equal $\sigma(\Lambda_A^{(1)})(\mathbf{x}, \xi)$. Moreover, for a real-valued function/distribution u , if $(\mathbf{x}, \xi) \in \text{WF}(u)$ then $(\mathbf{x}, -\xi) \in \text{WF}(u)$. This implies that $\mathbf{I}_B$, $\Lambda_A^{(1)}$, and $\mathbf{I}_K^{(1)}$ will produce similar migration outcomes.

4.5. Conclusion and recommendation. From our above explorations, we draw the following conclusion:

- For true-amplitude migration, we recommend using the classical migration formula from Example 2.2 (iii). This formula is straightforward to implement, efficient, and provides accurate results, as equality on the left in (24) applies accordingly. Recall that a needs to be calculated from (7).

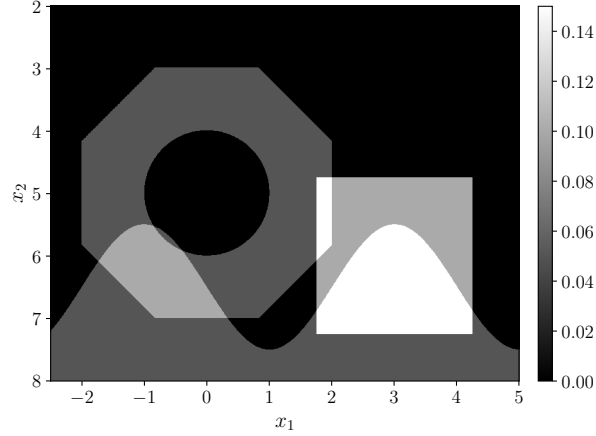


FIGURE 5. Perturbation n of the macro velocity $v(\mathbf{x}) = 0.1x_2 + 0.5$ to define ν_p in (29).

- If a is unknown or if the singularities of the ground truth are to be emphasized, we recommend employing a migration formula represented by one of the imaging operators $\Lambda_\theta^{(2)}$ (25) and $\tilde{\Gamma}_\theta^{(2)}$ (28), which are of order 1. Note that both migration formulas are based on newly discovered weight functions for the backprojection. These formulas effectively enhance singularities irrespective of their location because (23) and (27) hold accordingly for these operators.

5. NUMERICAL ILLUSTRATIONS

We illustrate our recommendations from Section 4.5 by numerical experiments using seismogram data g given by the right-hand side of (3), that is, we numerically solve the wave equation (1) for u and $\tilde{u}$.

In the first set of experiments (Section 5.1), we demonstrate the predicted mathematical properties of the various imaging operators $\mathcal{L}$ from (10) (Kirchhoff migration formulas). To this end, we selected a setting in which all the necessary assumptions were met, particularly the geometric optics assumption (see Section 2) and the Bolker condition (see the first two paragraphs of Section 4 and Remark 4.1). Recall that these assumptions impose restrictions on the macro velocity model. The second set of experiments (Section 5.2) involves a setting where these assumptions are presumably violated.

5.1. The clean mathematical setting. In view of (2), for the chosen macro velocity $v(\mathbf{x}) = 0.1x_2 + 0.5$, see (15), we set

$$(29) \quad \nu_p^2(\mathbf{x}) := \frac{v^2(\mathbf{x})}{1 + n(\mathbf{x})},$$

where n is displayed in Figure 5. The maximum amplitude of n is chosen small enough so that the linearization error from using the Born approximation does not completely corrupt g .

In (1) we replace the source term by a scaled, time-shifted, and normalized Gaussian pulse. The resulting equation is then solved on the computational domain $[-15, 20] \times [0, 15]$ using the wave solver of the open-source toolbox PySIT [7] with a step size 0.042 in each direction. To suppress spurious reflections, the computational domain is equipped with a perfectly matched layer (PML) absorbing boundary. With ν_p and v as inputs for the wave solver, we compute the

seismograms

$$u(t; \mathbf{x}_r(s), \mathbf{x}_s(s)) \quad \text{and} \quad \tilde{u}(t; \mathbf{x}_r(s), \mathbf{x}_s(s)) \quad \text{for} \quad (s, t) \in S \times T,$$

using the common offset configuration (14) with $\alpha = 5$. Further, S contains 125 points equidistantly distributed in $S = [-10, 15]$ and T contains 7200 equally spaced points in the time interval $[17.64, 42.637]$.

Approximating the integral on the right-hand side of (3) using the trapezoidal sum yields the seismogram data g . The different migration formulas acting on g have been implemented according to [14]; see [10] for the needed routines.

First, Figure 6 shows the results of true-amplitude migration in relation to formulas analyzed in previous sections. The upper right image depicts classical Kirchhoff migration, as seen in Example 2.2 (iii). The image directly below is the output of our proposed modification when the amplitude a is unavailable; see the left side of (28). The upper left image in Figure 6 illustrates the formula of Example 2.2 (v) for $i = 1$ and the image below results from this formula when modified to the situation with an unknown a ; see (25) ($i = 1$). All four images of Figure 6 are qualitatively similar and show the same wave fronts, which is consistent with our microlocal analysis. Furthermore, images on the same row are quantitatively close as well. One reason for this is certainly that the corresponding imaging operators share the same principal symbol.

Figure 7 illustrates two migration formulas that emphasize singularities and do not require a . These formulas originate from Examples 2.2 (v) for $i = 2$ and (iv); they have been adapted for the case of missing a , see (25) ($i = 2$) and (28) (right). As expected from our analysis, they yield very similar outcomes, both qualitatively and quantitatively.

In each of the simulated migration outcomes in Figures 6 and 7, there are artifact curves, such as the parenthesis-shaped curves around the square on the lower left. These are likely limited data artifacts as analytically described in [28]. For a discussion of these endpoint artifacts, as they are called in [3], from a geophysical point of view, we refer to Section 5.7 of [3].

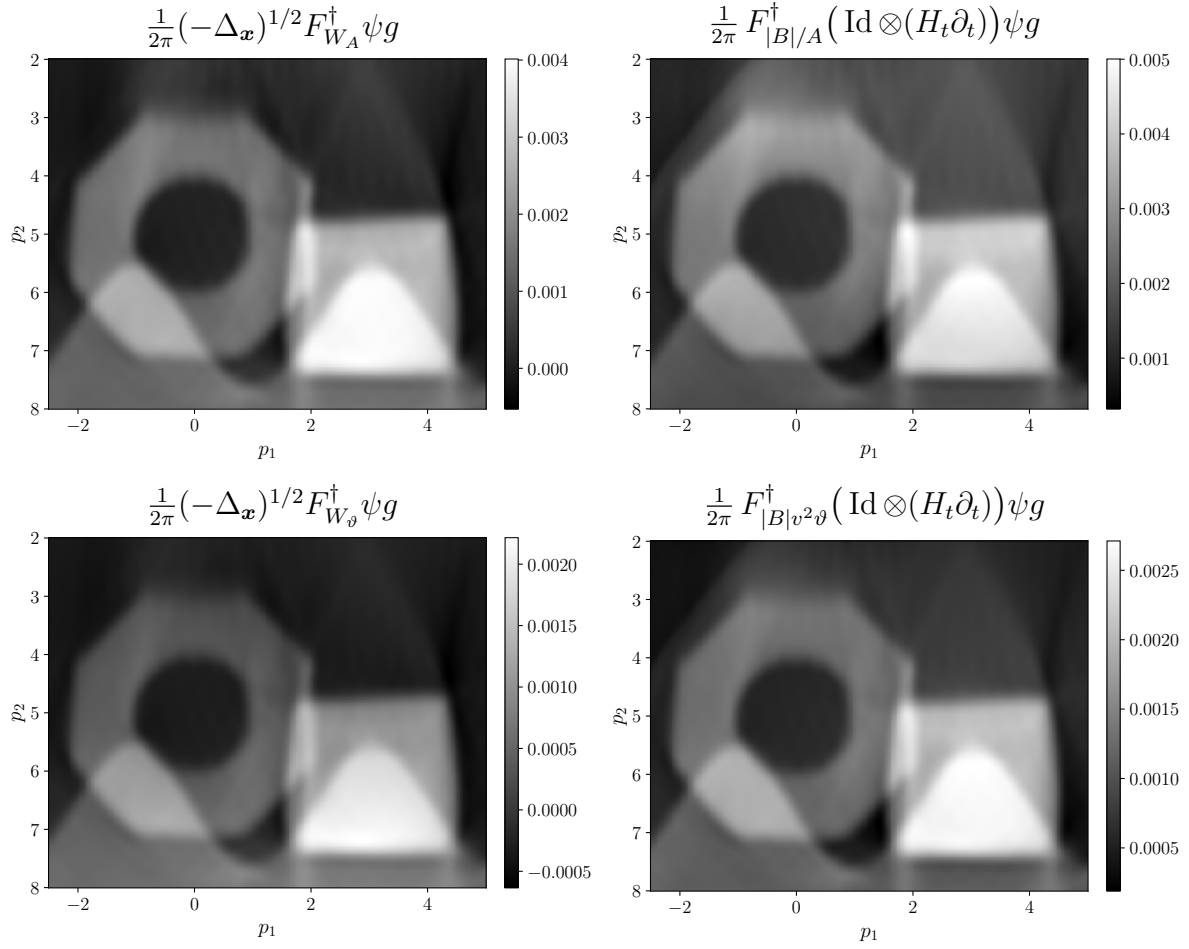


FIGURE 6. Results of true-amplitude migration. The migration formula is specified at the top of each image. Top row: Amplitude a is known; bottom row: a is not known.

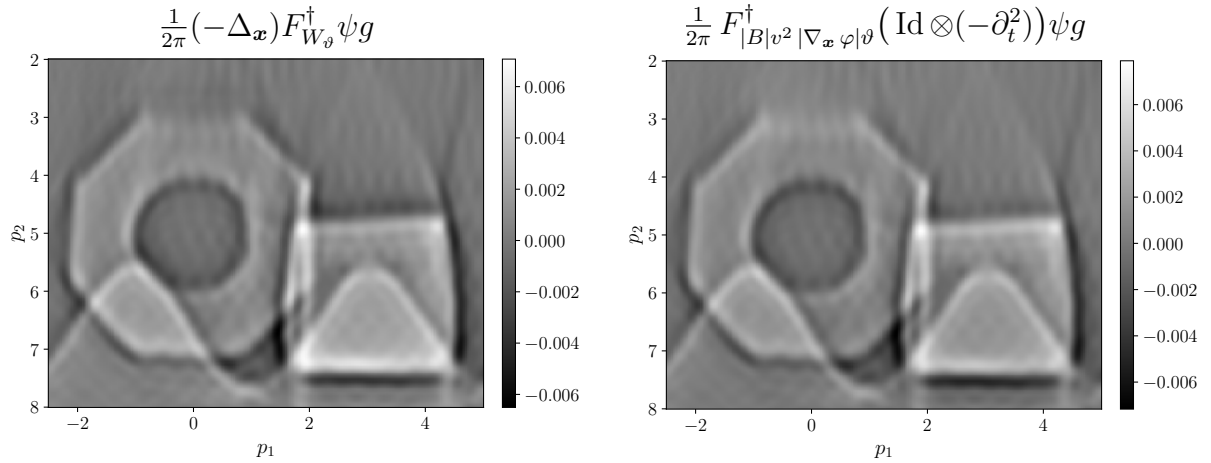


FIGURE 7. Results of migration that emphasizes singularities and do not require a . The migration formula is specified at the top of each image.

5.2. Violating the theoretical assumptions: the Marmousi model. Here, the true compression wave speed ν_p is given by the Marmousi model; see the top of Figure 3, which represents a 2D cut through a complex geological structure with strong velocity variations, making it a challenging test case. We consider two models for the macro velocity v : one that depends only on the depth variable, and one that depends on both the lateral and depth variables. In either case, according to (2), the ideal reconstruction outcome is $\mathcal{L}n$ where

$$n = \frac{v^2 - \nu_p^2}{\nu_p^2},$$

which represents the high frequency content of ν_p .

The two macro velocity models are

(1) **Piecewise affine linear:** Here,

$$v_1(\mathbf{x}) = \begin{cases} v_0 & : x_2 \leq 400\text{m}, \\ v_0 + \beta(x_2 - 400\text{m}) & : x_2 > 400\text{m}, \end{cases}$$

with $v_0 = 1500$ m/s and $\beta \approx 0.91 \text{ s}^{-1}$, see (15). This simple model approximates the general trend of increasing velocity with depth, but it ignores lateral variations. Please note that 400m is the depth of the seafloor in our Marmousi model.

(2) **Smoothed Marmousi:** A spatially smoothed version v_2 of ν_p , obtained by convolving ν_p with a tensor product Gaussian kernel having zero mean and standard deviations of 230m and 74m in the horizontal and vertical directions, respectively. This model preserves the large-scale velocity structure while removing the fine-scale reflectivities.

Figure 8 displays both macro velocities.

While the implementation from [10] is well suited for migration based on v_1 , it needs to be extended to handle macro velocities that depend on depth and lateral variables, as does v_2 . The underlying method and its implementation are detailed in Appendix C, where also the used parameter settings for our experiments can be found; see Table 1. The corresponding code is published in [12].

As in the previous subsection, we solve the wave equation, this time using the software Deepwave [30], to obtain

$$(30) \quad u(t; \mathbf{x}_r(s), \mathbf{x}_s(s)) \quad \text{and} \quad \tilde{u}(t; \mathbf{x}_r(s), \mathbf{x}_s(s)) \quad \text{for} \quad (s, t) \in S \times T$$

in the common offset configuration (14). The used offset α as well as the discrete sets S and T are specified in Table 1. Once more, the seismogram data g are simulated by applying the trapezoidal rule to the right-hand side of (3). Figure 9 displays the resulting seismograms.

5.2.1. Imaging operators of different orders. We apply migration formulas to the seismograms in Figure 9 via Algorithm 1 from Appendix C. These formulas are related to the imaging operators $\mathcal{I}_\theta^{(1)}$ and $\mathcal{I}_\theta^{(2)}$, see (28) and (25), respectively. Figures 10 and 11 show the migration outcomes. It is not surprising that the results based on v_2 are more accurate, especially in larger depths.

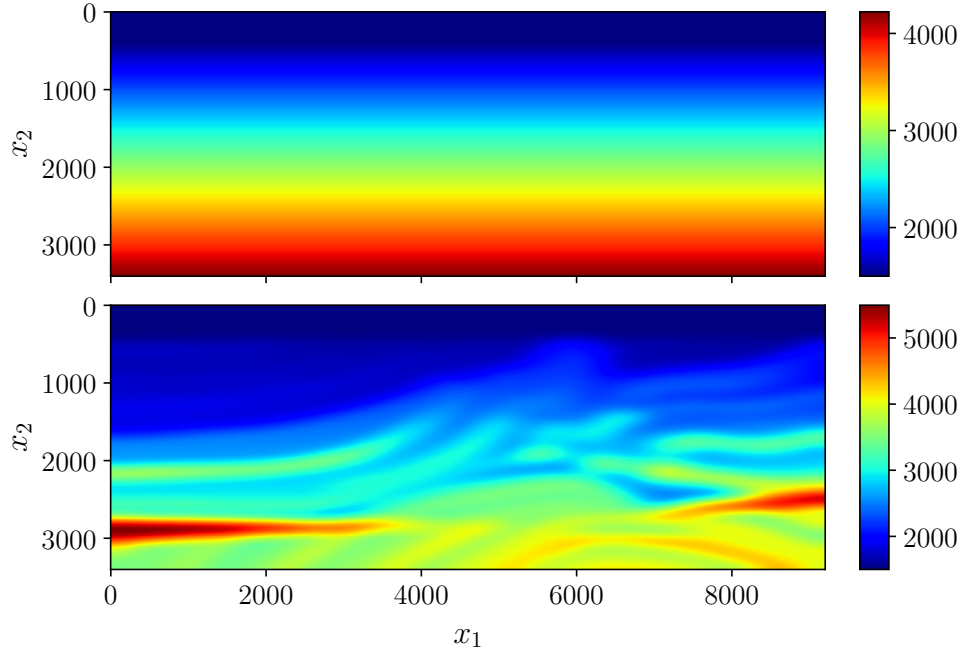


FIGURE 8. Macro velocity models in m/s. Top: linearly increasing velocity v_1 . Bottom: smooth version v_2 of ν_p .

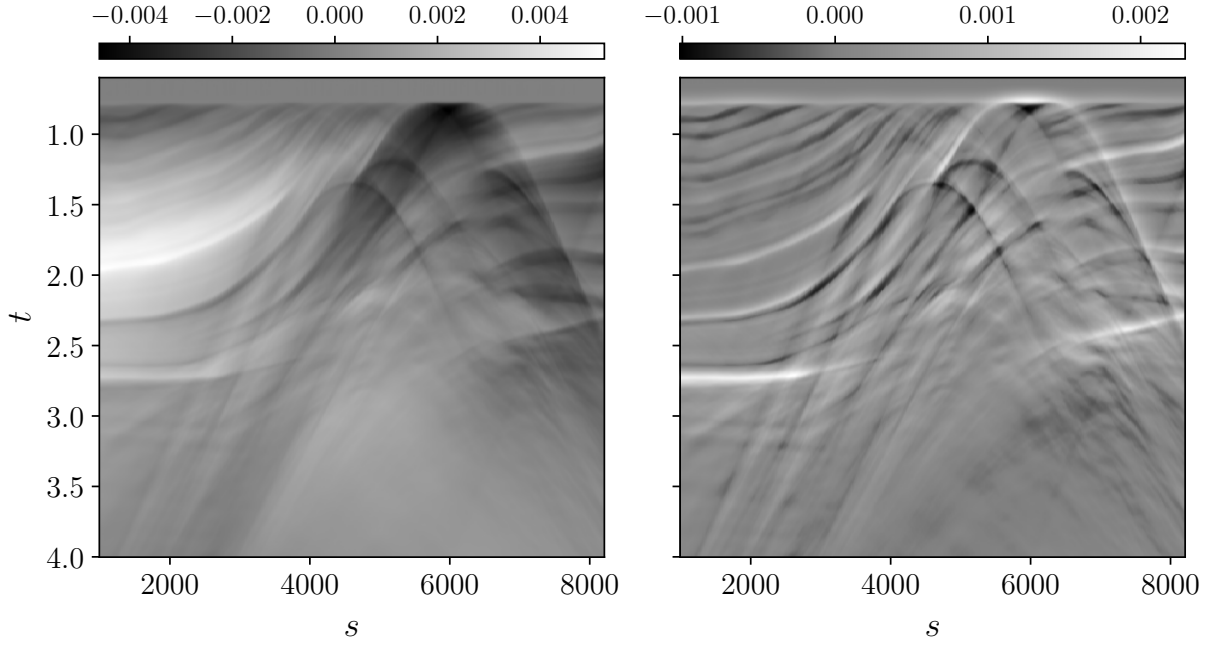


FIGURE 9. Common offset data g for the linear velocity model v_1 (left) and for the smooth model v_2 (right).

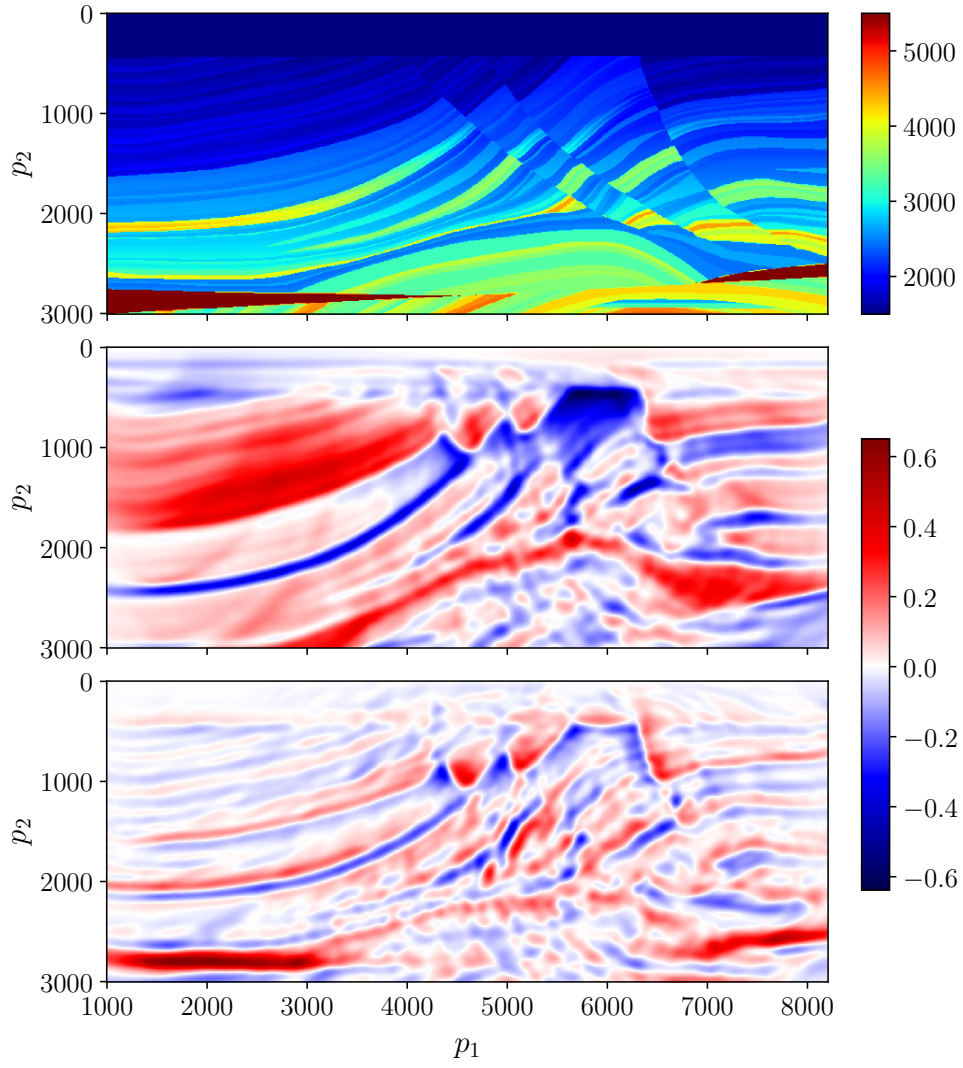


FIGURE 10. Results of migration formula $\frac{1}{2\pi} F_{|B|v^2\vartheta}^\dagger (\text{Id} \otimes (H_t \partial_t)) \psi g$. The corresponding imaging operator $I_\vartheta^{(1)}$ has order 0. Top: Marmousi ground truth. Middle: seismogram data g with respect to macro velocity v_1 . Bottom: seismogram data g with respect to macro velocity v_2 .

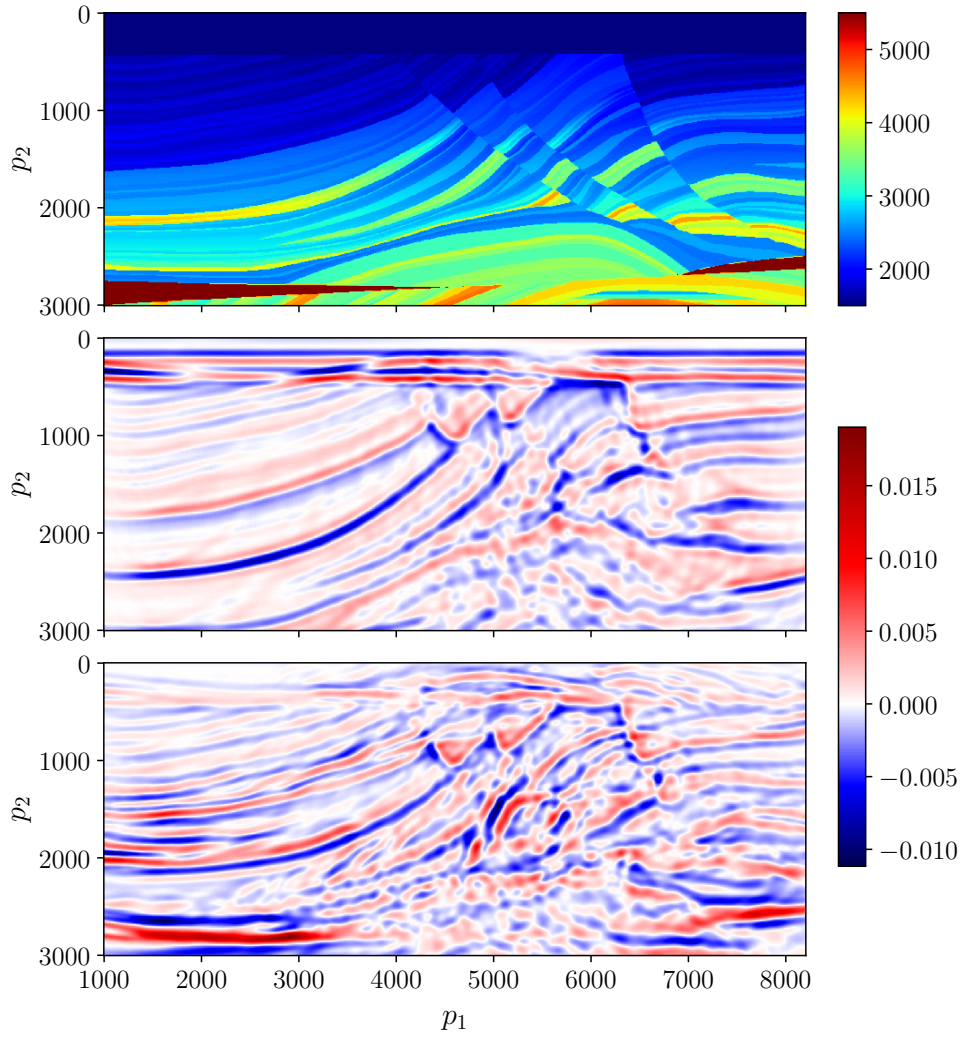


FIGURE 11. Results of migration formula $\frac{1}{2\pi}(-\Delta_{\mathbf{x}})F_{W_\vartheta}^\dagger\psi g$. The corresponding imaging operator $\Lambda_\vartheta^{(2)}$ has order 1. Top: Marmousi ground truth. Middle: seismogram data g with respect to macro velocity v_1 . Bottom: seismogram data g with respect to macro velocity v_2 .

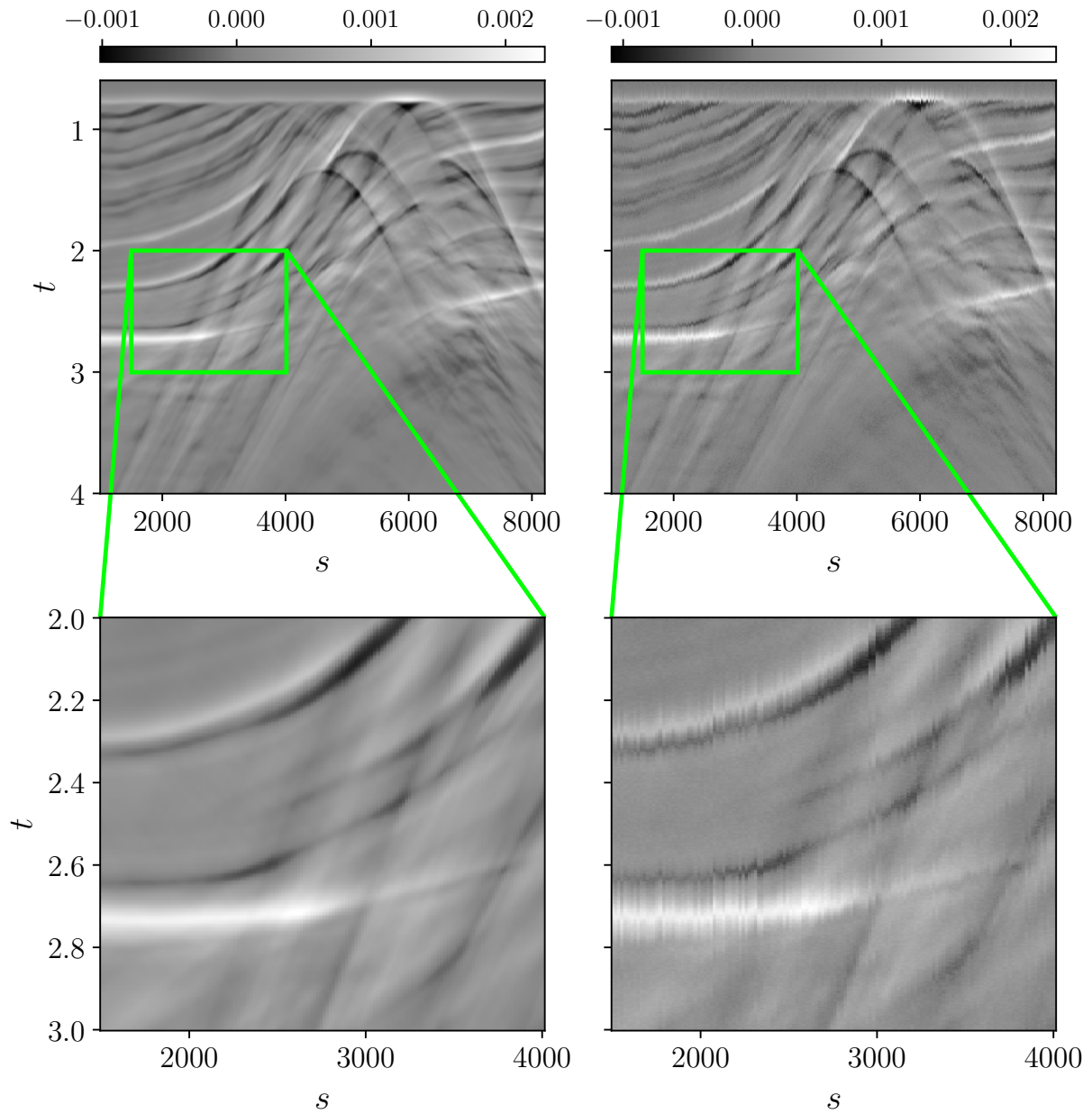


FIGURE 12. Seismogram data from the wave solver with respect to the macro velocity v_2 . Top left: g (agrees with the seismogram on the right of Figure 9). Top Right: perturbed seismogram g^ε from (31) (uncertainty radius 20m and $\varepsilon = 10\%$ relative noise). Bottom: zoom in of g and g^ε on an area as indicated.

5.2.2. A more realistic scenario. For a more realistic setting, we include additional uncertainties. First, the source and receiver positions used in the migration formulas are perturbed for computing the seismograms. To this end, we place the sources and receivers randomly uniformly distributed in the positive half ball ($x_2 \geq 0$) centered at the respective exact positions with a radius of 20m. The resulting perturbed source/receiver positions are indicated by $\overline{\mathbf{x}_s(s)}$ and $\overline{\mathbf{x}_r(s)}$, respectively. Thus, we generate the seismogram data $\overline{g}$ as above, however, using the wave fields

$$u(t; \overline{\mathbf{x}_r(s)}, \overline{\mathbf{x}_s(s)}) \quad \text{and} \quad \tilde{u}(t; \overline{\mathbf{x}_r(s)}, \overline{\mathbf{x}_s(s)}) \quad \text{for} \quad (s, t) \in \mathbf{S} \times \mathbf{T}$$

rather than the wave fields from (30).

Second, we add measurement noise to these seismograms according to

$$(31) \quad g^\varepsilon = \overline{g} + \varepsilon \|\overline{g}\|_* \frac{\mathbf{N}}{\|\mathbf{N}\|_*}, \quad \varepsilon > 0,$$

where $\mathbf{N}$ is an array of the dimension of $\overline{g}$ containing uniformly distributed random numbers in $[-1, 1]$ and ε is the relative noise level with respect to the Frobenius norm $\|\cdot\|_*$.

The perturbed seismogram data g^ε , which we use in the remainder of this section, is based on the macro velocity model v_2 . It can be seen in the top and bottom right of Figure 12. Migration results from this g^ε are presented in Figures 13 and 14, where they are also compared to the corresponding outputs with respect to g from the left of Figure 12. It turns out that both migration formulas are rather robust against perturbations in the source and receiver positions, as well as against measurement noise (see [22] for a theoretical argument of this observation). We are convinced that this remarkable robustness accounts for their success in geophysical applications.

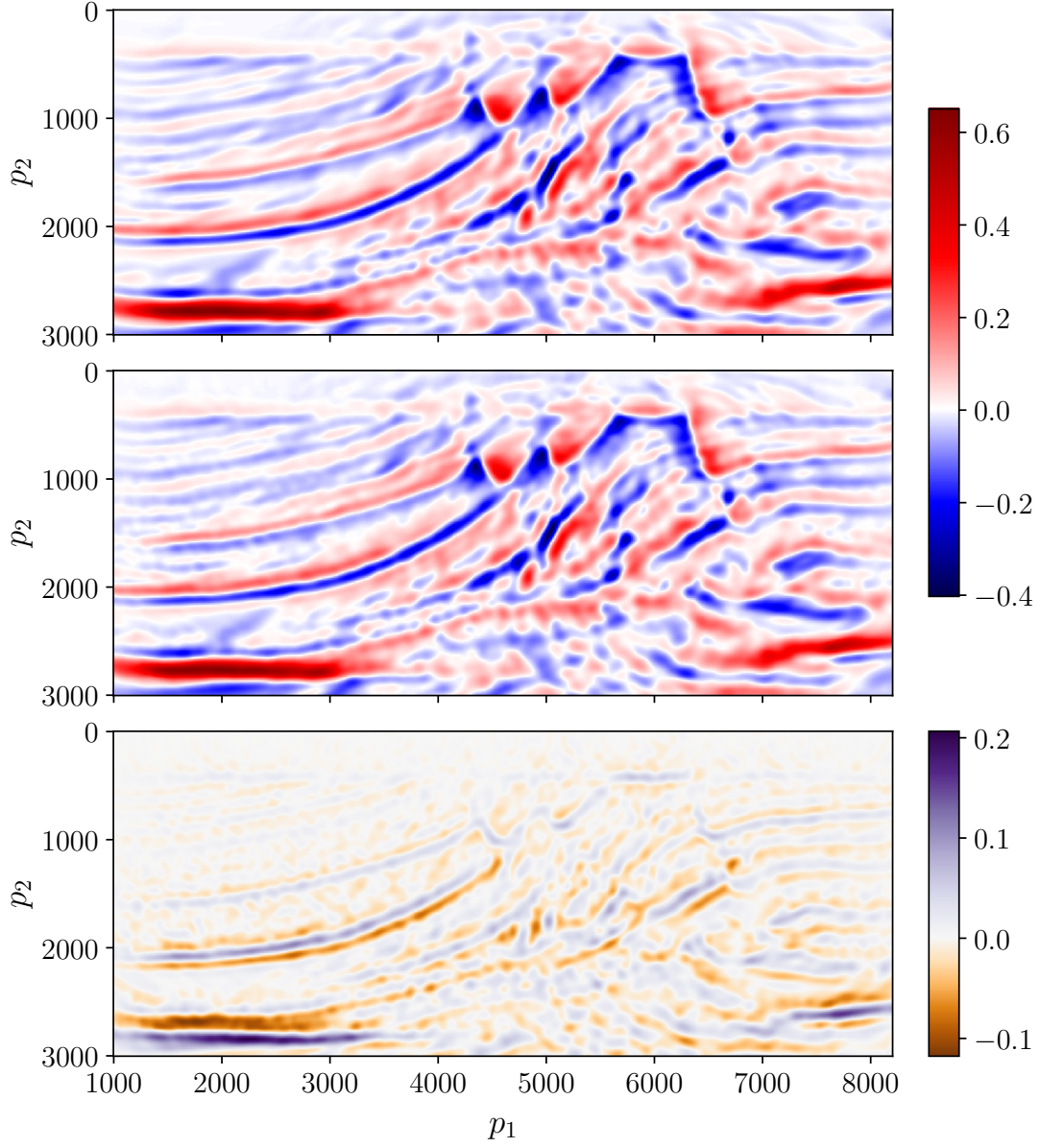


FIGURE 13. Migration based on the imaging operator $I_\vartheta^{(1)}$ of order 0 using the macro velocity model v_2 .

Top: $\frac{1}{2\pi} F_{|B|v^2\vartheta}^\dagger (\text{Id} \otimes (H_t \partial_t)) \psi g$ with g from Figure 12 (left).

Middle: $\frac{1}{2\pi} F_{|B|v^2\vartheta}^\dagger (\text{Id} \otimes (H_t \partial_t)) \psi g^\varepsilon$ with g^ε from Figure 12 (right).

Bottom: difference between results from the top and the middle.

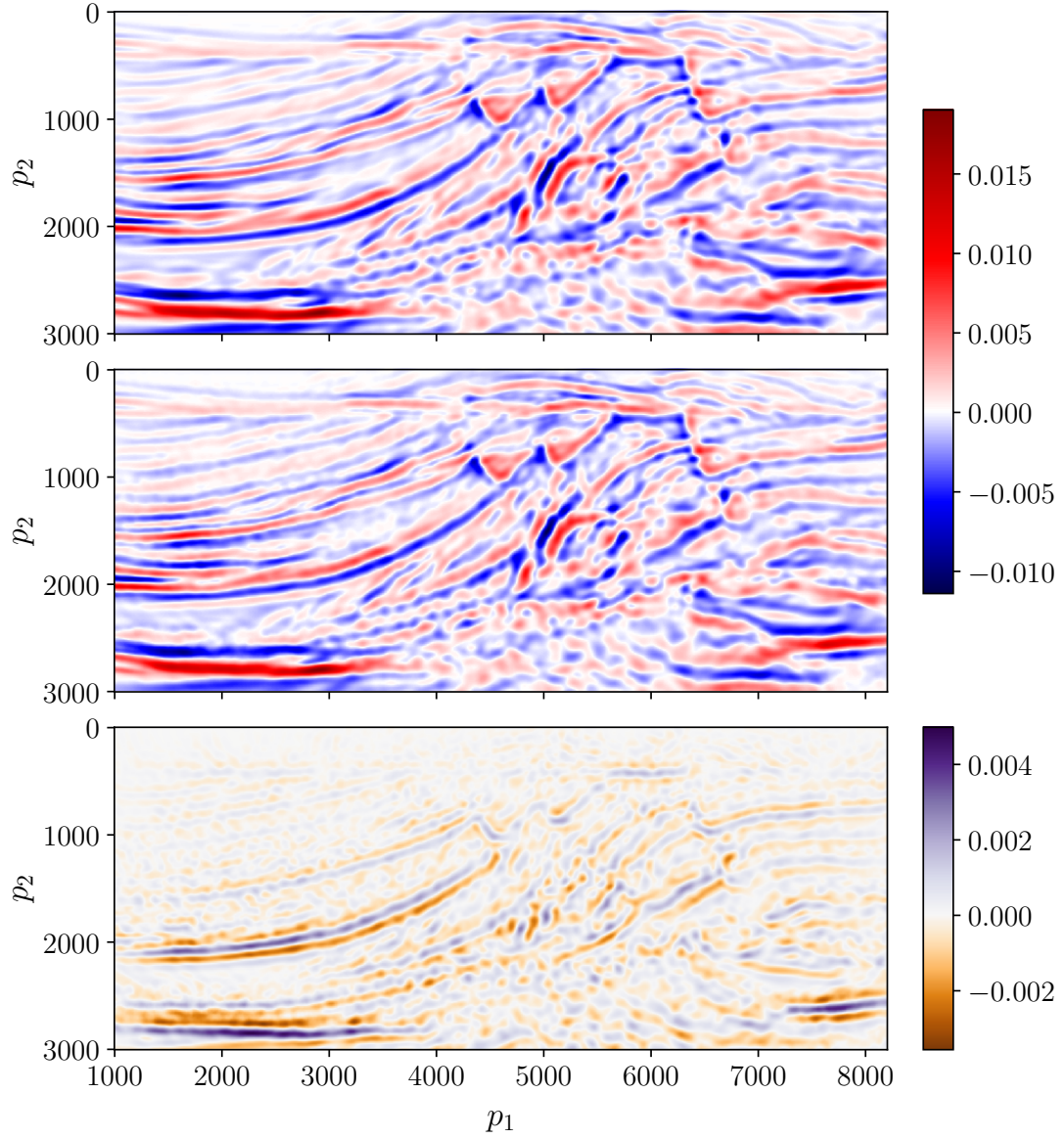


FIGURE 14. Migration based on the imaging operator $\Lambda_\vartheta^{(2)}$ of order 1 using the macro velocity model v_2 .

Top: $\frac{1}{2\pi}(-\Delta_{\mathbf{x}})F_{W_\vartheta}^\dagger\psi g$ with g from Figure 12 (left).

Middle: $\frac{1}{2\pi}(-\Delta_{\mathbf{x}})F_{W_\vartheta}^\dagger\psi g^\varepsilon$ with g^ε from Figure 12 (right).

Bottom: difference between results from the top and the middle.

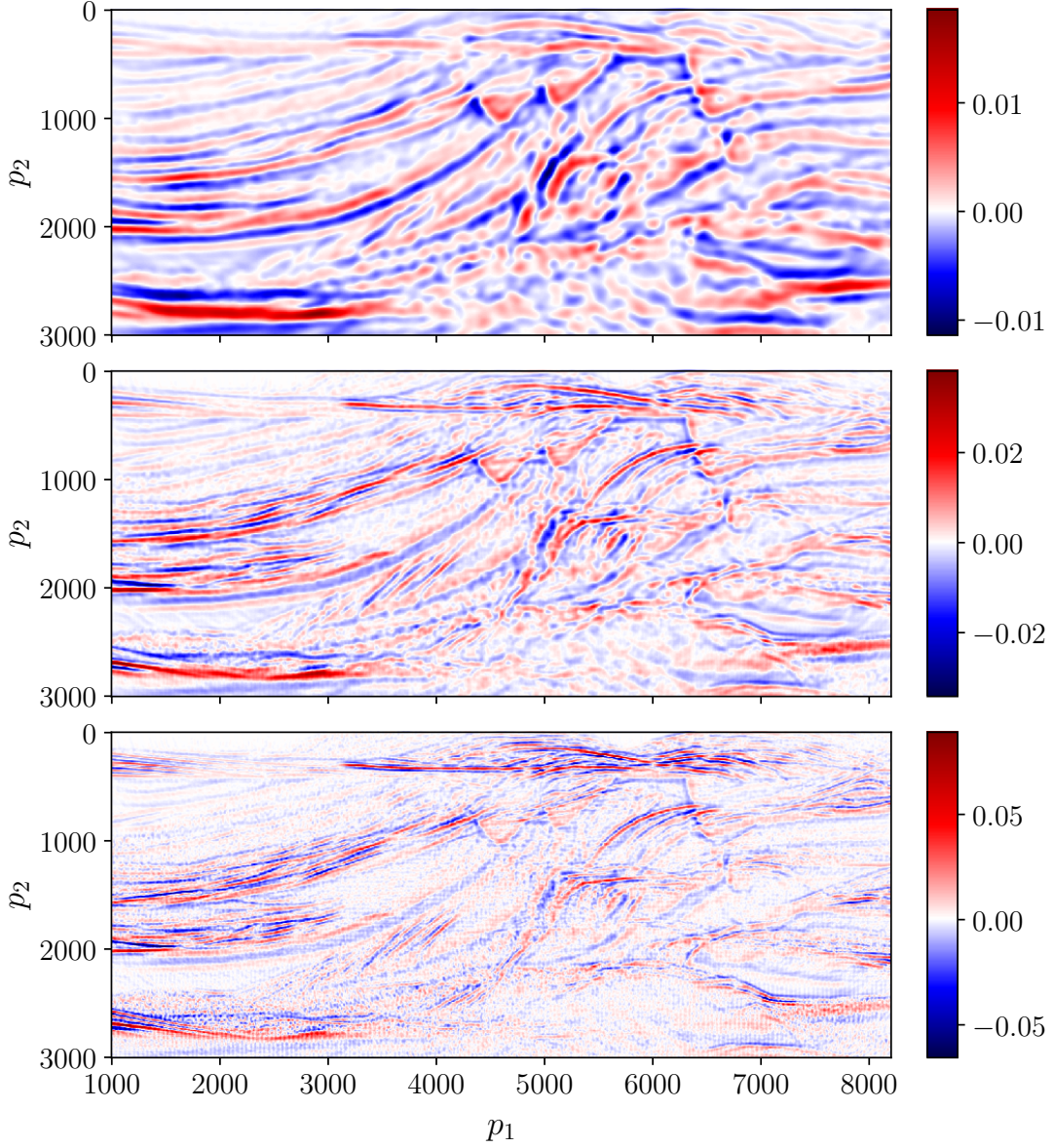


FIGURE 15. Comparison of migration using different regularization parameter γ under perturbed data g^ε from (31). The underlying imaging operator is $\Lambda_\theta^{(2)}$. Top: $\gamma = 100$. Middle: $\gamma = 50$. Bottom: $\gamma = 30$.

5.2.3. The regularizing effect. In the final experiment, we demonstrate the regularizing effect of the parameter γ , an essential ingredient in our implementation of the migration formulas; see Appendix C.1. Until now, we have used $\gamma = 100\text{m}$ throughout. At the top of Figure 15 you see $\frac{1}{2\pi}(-\Delta_x)F_{W_\theta}^\dagger\psi g^\varepsilon$ with g^ε from (31) approximated by Algorithm 1 with $\gamma = 100\text{m}$ (same as the middle image of Figure 14). The outputs for $\gamma = 50\text{m}$ and $\gamma = 30\text{m}$, where all other quantities are frozen, are shown in the middle and bottom of Figure 15, respectively. With γ decreasing, finer structures become visible, while the noise in the data begins to corrupt the migration outcome. In particular, coherent, elongated structures disintegrate.

We would like to point out that, in principle, the value of γ can be selected individually for each point in the reconstruction grid.

Conflict of interest. All authors declare no conflicts of interest.

Data availability statement. Our software, which we used to execute the numerical experiments in Sections 5.1 and 5.2, is freely available from [10] and [12], respectively.

APPENDIX A. NORMALIZATION OF THE AMPLITUDE FUNCTION

In this appendix, we explain the normalization of the amplitude function in the transport equation (7).

Recall that the spatial dimension is 2. To derive the relation (3) with F given in (4) from the ansatz (2), one relies on the progressing wave representation

$$(32) \quad \tilde{u}(t; \mathbf{x}, \mathbf{x}_s) \approx a(\mathbf{x}, \mathbf{x}_s) \Psi(t - \tau(\mathbf{x}, \mathbf{x}_s)), \quad \Psi(t) := \begin{cases} t^{-1/2} & : t > 0, \\ 0 & : t \leq 0, \end{cases}$$

for the reference solution $\tilde{u}$, which is the solution of the wave equation (1) with the macro velocity v , i.e., $\nu_p = v$. This representation of $\tilde{u}$ is qualitatively accurate if τ and a solve the eikonal and transport equation,

$$|\nabla_{\mathbf{x}} \tau(\cdot, \mathbf{y})|^2 = v^{-2}(\cdot) \quad \text{and} \quad \operatorname{div}_{\mathbf{x}}(a^2(\cdot, \mathbf{y}) \nabla_{\mathbf{x}} \tau(\cdot, \mathbf{y})) = 0,$$

respectively, see [34] and also [6, 9]. At this point no normalization of a is required.

In the case of a constant sound speed, $v = c$, the reference solution $\tilde{u}$ is the fundamental solution of the wave equation:

$$(33) \quad \tilde{u}(t; \mathbf{x}, \mathbf{x}_s) = \frac{1}{2\pi \sqrt{t + |\mathbf{x} - \mathbf{x}_s|/c}} \Psi(t - |\mathbf{x} - \mathbf{x}_s|/c),$$

see, e.g., [9, Appendix to Chapter 5]. When the wave front reaches the point $\mathbf{x}$ at time $t = |\mathbf{x} - \mathbf{x}_s|/c$, the factor preceding Ψ in (33) becomes

$$(34) \quad \frac{1}{2\pi} \sqrt{\frac{c}{2|\mathbf{x} - \mathbf{x}_s|}}.$$

Therefore, a in (32) should behave the same in the limit at the source point ($\mathbf{x} \rightarrow \mathbf{x}_s$), even in a heterogeneous medium, where c has to be replaced by $v(\mathbf{x})$. This physically meaningful normalization is achieved through the prescribed limit in (7).

Remark A.1. Here we comment briefly on the normalization in three dimensions where the progressing wave ansatz (32) requires us to set $\Psi := \delta$, since the fundamental solution for constant sound speed c is

$$\tilde{u}(t; \mathbf{x}, \mathbf{x}_s) = \frac{1}{4\pi t} \delta(ct - |\mathbf{x} - \mathbf{x}_s|) = \frac{1}{4\pi ct} \delta(t - |\mathbf{x} - \mathbf{x}_s|/c), \quad t \geq 0.$$

As above, we find the normalization $\lim_{\mathbf{x} \rightarrow \mathbf{y}} a(\mathbf{x}, \mathbf{y}) |\mathbf{x} - \mathbf{y}| = 1/(4\pi)$ to be used in (7).

APPENDIX B. TEMPERED DISTRIBUTIONS

Tempered distributions are generalized functions for which the Fourier transform can be defined. For $\phi \in \mathcal{C}^\infty(\mathbb{R}^d)$ and $k \in \mathbb{N}_0$ set

$$p_k(\phi) := \sup_{\mathbf{x} \in \mathbb{R}^d, |\alpha| \leq k} (1 + |\mathbf{x}|)^k |\partial^\alpha \phi(\mathbf{x})|.$$

The space $\mathcal{S}$ of *tempered* or *rapidly decreasing* functions is given by

$$\mathcal{S} := \{ \phi \in \mathcal{C}^\infty(\mathbb{R}^d) : \forall k \in \mathbb{N}_0 : p_k(\phi) < \infty \}.$$

The functions in $\mathcal{S}$ and their derivatives are integrable, and they decrease faster to zero at infinity than any polynomial can grow: Let $\phi \in \mathcal{S}$ then $P \partial^\alpha \phi \in \mathcal{S}$ for any $\alpha \in \mathbb{N}_0^d$ and any polynomial P in d variables. We have that $e^{-|\cdot|^2/2} \in \mathcal{S}$. Further, smooth functions with compact support are in $\mathcal{S}$: $\mathcal{C}_0^\infty(\mathbb{R}^d) \subset \mathcal{S}$.

Definition B.1. A linear mapping $T: \mathcal{S} \rightarrow \mathbb{C}$ is a **tempered distribution** if there are $C_T > 0$ and $m \in \mathbb{N}_0$ such that $|T\phi| \leq C_T p_m(\phi)$ for all $\phi \in \mathcal{S}$. The space of tempered distributions is denoted by $\mathcal{S}'$ (dual space of $\mathcal{S}$).

The Dirac distribution $\delta_x(\phi) := \phi(x)$ is tempered. Further, all locally integrable functions $f: \mathbb{R}^d \rightarrow \mathbb{C}$ that grow at most polynomially at infinity are tempered distributions via the setting

$$(35) \quad T_f \phi := \int_{\mathbb{R}^d} f(x) \phi(x) dx$$

and identifying f with T_f .

To define the Fourier transform for tempered distributions, we note that $\mathcal{FS} \subset \mathcal{S}$ for $\mathcal{F}$ from (11). Moreover, for each $k \in \mathbb{N}_0$ there are a $k' \in \mathbb{N}_0$ and a $C_{\mathcal{F}} > 0$ such that

$$p_k(\widehat{\phi}) \leq C_{\mathcal{F}} p_{k'}(\phi) \quad \text{for all } \phi \in \mathcal{S}.$$

For integrable functions f and g we recall Parseval's relation

$$\int_{\mathbb{R}^d} \widehat{f}(x) g(x) dx = \int_{\mathbb{R}^d} f(y) \widehat{g}(y) dy$$

which suggests, with a view to (35), the following definition.

Definition B.2. The Fourier transform $\widehat{T}$ of the tempered distribution T is given by

$$\widehat{T}\phi := T\widehat{\phi} \quad \text{for all } \phi \in \mathcal{S}.$$

We stress the fact that $\widehat{T} \in \mathcal{S}'$: Obviously, $\widehat{T} = T \circ \mathcal{F}$ is linear as a composition of two linear mappings and

$$|\widehat{T}\phi| = |T\widehat{\phi}| \leq C_T p_m(\widehat{\phi}) \leq C_T C_{\mathcal{F}} p_{m'}(\phi) \quad \text{for all } \phi \in \mathcal{S}.$$

For a function f that is identified with the tempered distribution T_f , we set $\widehat{f} = \widehat{T}_f$.

As an example, let us compute $\widehat{\delta}_x$. For $\phi \in \mathcal{S}$ we have that

$$(36) \quad \widehat{\delta}_x(\phi) = \delta_x(\widehat{\phi}) = \widehat{\phi}(x) = \int_{\mathbb{R}^d} e^{-i x \cdot \xi} \phi(\xi) d\xi = T_{e^{-i x \cdot \cdot}} \phi.$$

Hence, $\widehat{\delta}_x(\cdot) = e^{-i x \cdot \cdot}$. Especially, $\widehat{\delta}_0 = 1$.

We calculate $\widehat{1}$ as another example:

$$\widehat{1}(\phi) = \widehat{T}_1(\phi) = T_1(\widehat{\phi}) = \int_{\mathbb{R}^d} \widehat{\phi}(\xi) e^{i 0 \cdot \xi} d\xi = (2\pi)^d \phi(0) = (2\pi)^d \delta_0(\phi).$$

Hence, $\widehat{1} = (2\pi)^d \delta_0$.

Definition B.3. Let $T \in \mathcal{S}'$ and $\Omega \subset \mathbb{R}^d$ be open. We say T is zero in Ω if $T\phi = 0$ for any $\phi \in \mathcal{C}_0^\infty(\Omega)$. The **support** of T ($\text{supp } T$) is the smallest closed subset outside of which $T = 0$, so this means that $T = 0$ in $\mathbb{R}^d \setminus \text{supp } T$.

We have $\text{supp } \delta_x = \{x\}$.

APPENDIX C. NUMERICAL REALIZATION BY APPROXIMATE INVERSE

C.1. The concept of approximate inverse. Let $\mathcal{L} = KF_W^\dagger(\text{Id} \otimes L)\psi F$ be the imaging operator from (10) and let $\mathcal{M}_p \subset X$ be the reconstruction mesh. Our numerical scheme to reconstruct $\mathcal{L}n(\mathbf{p})$ for all $\mathbf{p} \in \mathcal{M}_p$ from seismogram data is based on the concept of approximate inverse [24, 33] where one wants to compute a smoothed version of $\mathcal{L}n(\mathbf{p})$. To this end, we consider the convolution

$$\mathcal{L}n(\mathbf{p}) \approx \mathcal{L}n \star e_\gamma(\mathbf{p}) = \int_X \mathcal{L}n(\mathbf{x}) e_\gamma(\mathbf{x} - \mathbf{p}) d\mathbf{x},$$

where e_γ is a suitable mollifier, that is, a function depending on a parameter $\gamma > 0$ such that $\int_X e_\gamma(\mathbf{x}) d\mathbf{x} = 1$ for all $\gamma > 0$ and $T_{e_\gamma}\phi \rightarrow \phi(\mathbf{0}) = \delta_0(\phi)$ as $\gamma \rightarrow 0$ for any $\phi \in \mathcal{S}$, see (35) for T_{e_γ} .⁸ Thus, $\mathcal{L}n \star e_\gamma(\mathbf{p}) \rightarrow \mathcal{L}n(\mathbf{p})$ and γ acts as a *regularization* parameter. In our computations in Section 5 we used the compactly supported mollifier

$$e_\gamma(\mathbf{x}) = \frac{4}{\pi \gamma^8} \begin{cases} (\gamma^2 - |\mathbf{x}|^2)^3 : |\mathbf{x}| < \gamma, \\ 0 : |\mathbf{x}| \geq \gamma. \end{cases}$$

For any mollifier, it can be shown [14] that

$$(37) \quad \mathcal{L}n \star e_\gamma(\mathbf{p}) = \int_S \int_0^T (\psi F n)(s, t) v_{\mathbf{p}, \gamma, W}(s, t) dt ds,$$

where

$$v_{\mathbf{p}, \gamma, W} := (\text{Id} \otimes L^*) \tilde{v}_{\mathbf{p}, \gamma, W}, \quad \tilde{v}_{\mathbf{p}, \gamma, W} := F_W K^* e_\gamma(\cdot - \mathbf{p}),$$

$$F_W n(s, t) := \int_{\mathcal{R}(s, t)} \frac{n(\mathbf{x}) W(s, \mathbf{x})}{|\nabla_{\mathbf{x}} \varphi(s, \mathbf{x})|} d\mathbf{s}(\mathbf{x}).$$

The operator F_W is the generalized version of (4), while $\mathcal{R}(s, t)$ are the reflection isochrones as in (8). Moreover, the superscript $*$ indicates the formal L^2 -adjoint of the respective operator.

Now, assume that we measure the seismogram data at discrete times $T = \{t_1, \dots, t_{n_T}\} \subset [0, T]$ and at discrete points $S = \{s_1, \dots, s_{n_S}\}$ from the aperture S . For the ease of presentation and for our computations in Section 5.2, we restrict ourselves to equidistant sampling with $h_S = s_{i+1} - s_i$ and $h_T = t_{j+1} - t_j$. This allows us to use the simplified trapezoidal rule to approximate corresponding integrals. Extension to non-uniform grids requires only replacing the simplified trapezoidal rule with its generalized counterpart. Applying the trapezoidal rule to the right of (37) and replacing $F n$ with the discrete seismogram data g leaves us with the discrete migration formula

$$(38) \quad KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{p}) \approx h_S h_T \sum_{i=1}^{n_S} \sum_{j=1}^{n_T} (\psi g)(s_i, t_j) v_{\mathbf{p}, \gamma, W}(s_i, t_j).$$

The functions $\{v_{\mathbf{p}, \gamma, W}\}_{\mathbf{p} \in \mathcal{M}_p}$ are called reconstruction kernels. They are independent of the measurements g and must be computed numerically, which can be done very efficiently when the macro velocity model only depends on depth; see [14, 10]. In the remainder of this appendix, we present an efficient implementation and parallelization of the right side of (38) for arbitrary macro velocity models.

⁸In the sense of distributions we have the convergence $e_\gamma \rightarrow \delta_0$ as $\gamma \rightarrow 0$.

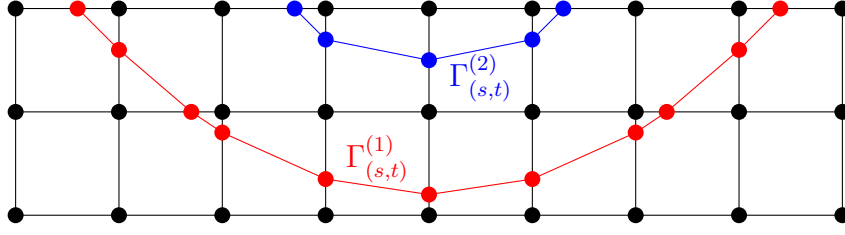


FIGURE 16. Isochrone contours (colored lines) extracted from the discrete travel time field (black grid) using Marching Squares for one fixed pair $(s, t) \in \mathcal{S} \times \mathcal{T}$. Contour vertices (colored dots) are located at subgrid positions via linear interpolation. Note that both colored lines belong to the same contour $\mathcal{R}(s, t)$. One contour may consist of multiple segments.

C.2. Numerical computation of reconstruction kernels. Let $\mathcal{M}_X$ be a computational mesh covering the domain of interest X , and containing the reconstruction mesh $\mathcal{M}_p$. For fixed $(s, t) \in \mathcal{S} \times \mathcal{T}$, the reflection isochrone $\mathcal{R}(s, t)$ is approximated using the Marching Squares algorithm [23] applied to the discrete travel time field $\varphi_h = \varphi(s, \mathcal{M}_X)$. This yields a collection of piecewise linear contours $\{\Gamma_{(s,t)}^{(m)}\}_{m=1}^{M(s,t)}$, where each $\Gamma_{(s,t)}^{(m)} = \{\mathbf{x}_k^{(m)}\}_{k=1}^{N_{(s,t)}^{(m)}}$ consists of $N_{(s,t)}^{(m)}$ vertices at subgrid positions (see Figure 16).

The integral defining $\tilde{v}_{p,\gamma,W}(s, t)$ is then approximated by numerical quadrature along all contours,

$$(39) \quad \tilde{v}_{p,\gamma,W}(s, t) \approx \sum_{m=1}^{M(s,t)} \sum_{k=1}^{N_{(s,t)}^{(m)}} w_k^{(m)} \frac{W(s, \mathbf{x}_k^{(m)})}{|\nabla \varphi(s, \mathbf{x}_k^{(m)})|} (K^* e_\gamma)(\mathbf{x}_k^{(m)} - \mathbf{p}),$$

where w_k denotes the quadrature weight for the trapezoidal rule:

$$(40) \quad w_k^{(m)} = \frac{1}{2} \cdot \begin{cases} |\mathbf{x}_2 - \mathbf{x}_1|, & k = 1, \\ |\mathbf{x}_{N_{(s,t)}} - \mathbf{x}_{N_{(s,t)}-1}|, & k = N_{(s,t)}, \\ |\mathbf{x}_{k+1} - \mathbf{x}_k| + |\mathbf{x}_k - \mathbf{x}_{k-1}|, & \text{otherwise.} \end{cases}$$

Travel time computation. The travel time field $\varphi(s, \cdot) = \tau(\mathbf{x}_s, \cdot) + \tau(\mathbf{x}_r, \cdot)$ is obtained by solving the eikonal equation (6) separately for source $\mathbf{x}_s = (s - \alpha, 0)$ and receiver $\mathbf{x}_r = (s + \alpha, 0)$ using the factored Fast Marching Method [25].

The acquisition geometry and discretization parameters for the Marmousi experiments in Section 5.2 are summarized in Table 1.

C.3. Reconstruction algorithm. A naive implementation of the formula in (37) would proceed pointwise: for each $\mathbf{p} \in \mathcal{M}_p$, compute $v_{p,\gamma,W}(s_i, t_j)$ for all $(s_i, t_j) \in \mathcal{S} \times \mathcal{T}$ and accumulate the weighted sum. However, this approach is unfeasible for two reasons:

- (1) The discrete travel time field $\varphi_h = \varphi(s, \mathcal{M}_X)$ must be recomputed for each $\mathbf{p}$, even though it depends only on s .
- (2) Storing the full kernel tensor $v_{p,\gamma,W}(s_i, t_j)$ for all $\mathbf{p}, s_i, t_j$ requires $\mathcal{O}(n_{\mathcal{M}_p} n_{\mathcal{S}} n_{\mathcal{T}})$ memory capacity, which becomes impractical for realistic problem sizes (e.g., $n_{\mathcal{M}_p} = 501^2$, $n_{\mathcal{S}} = n_{\mathcal{T}} = 1001$ requires ≈ 1.8 TB in double precision).

TABLE 1. Concrete parameter settings for the Marmousi experiments in Section 5.2.

Parameter	Value
$\mathcal{M}_X$	2301×851 equidistant points in $[0\text{m}, 9200\text{m}] \times [0\text{m}, 3400\text{m}]$
$\mathcal{M}_p$	1201×501 equidistant points in $[1000\text{m}, 8200\text{m}] \times [0\text{m}, 3000\text{m}]$
offset α	400m
S	$n_S = 401, s_1 = 1000\text{m}, s_{n_S} = 8200\text{m}$
T	$n_T = 1001, t_1 = 0.0003\text{s}, t_{n_T} = 4.0001\text{s}$
regularization param. γ	100m (unless stated otherwise)
tapering fct. ψ	$\psi(s, t) = 1$ for $s_1 + 250 \leq s \leq s_{n_S} - 250, t_1 \leq t \leq t_{n_T} - 0.2$. Otherwise smooth transition to 0 at the boundary of $[s_1, s_{n_S}] \times [t_1, t_{n_T}]$, see [14, App. A] for the specific function used here.

We therefore propose a *matrix-free* strategy. Writing out the discrete inner product explicitly:

$$(41) \quad KF_W^\dagger(\text{Id} \otimes L)g(\mathbf{p}) \approx h_S h_T \sum_{i=1}^{n_S} \left[\sum_{j=1}^{n_T} (\psi g)(s_i, t_j) \cdot \underbrace{(L^* \{ \tilde{v}_{\mathbf{p}, \gamma, W}(s_i, t_l) \}_{l=1}^{n_T})}_{\text{kernel time-series at fixed } s_i}(t_j) \right].$$

For *fixed* s_i , the quantity $\tilde{v}_{\mathbf{p}, \gamma, W}(s_i, t_j)$ can be evaluated efficiently for *all* $\mathbf{p}$ simultaneously:

- (1) Compute $\varphi_h = \varphi(s_i, \mathcal{M}_X)$ once using Fast Marching.
- (2) For each contour vertex $\mathbf{x}_k^{(m)} \in \Gamma_{(s_i, t_j)}^{(m)}$, evaluate $(K^* e_\gamma)(\mathbf{x}_k^{(m)} - \mathbf{p})$ for all $\mathbf{p}$.
- (3) Accumulate contributions weighted by $(\psi g)(s_i, t_j)$ directly into the reconstruction array.

It is important to note that, if $K^* e_\gamma$ is compactly supported, solely those $\mathbf{p}$ within this support around the current $\mathbf{x}_k^{(m)}$ will contribute, meaning we only need to consider these points in step (2). Concretely, we define $\text{supp}(\mathbf{x}_k^{(m)}) := \{\mathbf{p} \in \mathcal{M}_p : K^* e_\gamma(\mathbf{x}_k^{(m)} - \mathbf{p}) \neq 0\}$ as the region around the point $\mathbf{x}_k^{(m)}$ in which the term $K^* e_\gamma$ is non-zero. For instance, in the case of the operator $\Lambda_\partial^{(2)}$ from eq. (25), we have $\text{supp}(\mathbf{x}_k^{(m)}) = B_\gamma(\mathbf{x}_k^{(m)}) \cap \mathcal{M}_p$. We further use $n_{\text{supp}}(\mathbf{x}_k^{(m)}) := \#\text{supp}(\mathbf{x}_k^{(m)})$ to denote the cardinality of this set.

The *matrix-free* approach offers the following advantages:

- **Memory efficiency:** Only a single time-series per reconstruction point needs to be stored temporarily ($\mathcal{O}(n_{\mathcal{M}_p} n_T)$). For migration formulas with $L = \text{Id}$, we can even omit storing the time-series entirely, reducing the memory requirements down to $\mathcal{O}(n_{\mathcal{M}_p})$.
- **Computational efficiency:** The expensive travel time solve is performed only once per s_i .
- **Locality:** For compactly supported $K^* e_\gamma$, the inner loop iterates only over points $\mathbf{p} \in \text{supp}(\mathbf{x}_k^{(m)})$, reducing complexity drastically from $\mathcal{O}(n_{\mathcal{M}_p} N_{(s,t)})$ to $\mathcal{O}(n_{\text{supp}} N_{(s,t)})$ numerical operations. In the actual implementation, we use a bounding box containing $\text{supp}(\mathbf{x}_k^{(m)})$ to iterate over.

Algorithm 1 summarizes the overall procedure. Here, the discrete gradient $\nabla \varphi_h$ is computed via second-order finite differences (central differences in the interior, one-sided at boundaries).

Algorithm 1: Matrix-free implementation of the discrete migration formula (38)

Input: Seismogram data g
Output: Approximation to $KF_W^\dagger(\text{Id} \otimes L)g(\mathcal{M}_p)$

```

1 Apply cutoff:  $g \leftarrow \psi \cdot g$ 
2 Initialize result:  $\text{rec} \leftarrow \mathbf{0} \in \mathbb{R}^{n_{\mathcal{M}_p}}$ 
3 for  $i = 1, \dots, n_S$  do
4   Compute travel times:  $\varphi_h \leftarrow \tau(s_i - \alpha, \mathcal{M}_X) + \tau(s_i + \alpha, \mathcal{M}_X)$ 
5   Compute weight field:  $W_h \leftarrow W(s_i, \mathcal{M}_X)$ 
6   Compute integrand:  $I_h \leftarrow W_h / |\nabla \varphi_h|$ 
   /* Build kernel time-series for all  $p$  at fixed  $s_i$  */
7   Initialize kernel buffer:  $V \leftarrow \mathbf{0} \in \mathbb{R}^{n_{\mathcal{M}_p} \times n_T}$ 
8   for  $j = 1, \dots, n_T$  do
9     Extract contours:  $\{\Gamma_{(s_i, t_j)}^{(m)}\}_m \leftarrow \text{MarchingSquares}(\varphi_h, t_j)$ 
10    for  $m = 1, \dots, M_{(s_i, t_j)}$  do
11      for  $k = 1, \dots, N_{(s_i, t_j)}^{(m)}$  do
12        Let  $\mathbf{x}_k^{(m)}$  be the  $k$ -th vertex of  $\Gamma_{(s_i, t_j)}^{(m)}$ 
13        for  $p \in \text{supp}(\mathbf{x}_k^{(m)})$  do
14           $V_{p,j} \mathrel{+}= w_k^{(m)} \cdot \Pi I_h(\mathbf{x}_k^{(m)}) \cdot (K^* e_\gamma)(\mathbf{x}_k^{(m)} - p)$ 
15        end
16      end
17    end
18  end
   /* Apply time operator and accumulate */
19  for  $p \in \mathcal{M}_p$  do
20    Apply  $L^*$ :  $\tilde{V}_{p,\cdot} \leftarrow L^*(V_{p,1}, \dots, V_{p,n_T})$ 
21    Inner product:  $\text{rec}(p) \mathrel{+}= \sum_{j=1}^{n_T} g(s_i, t_j) \cdot \tilde{V}_{p,j}$ 
22  end
23 end
24 return  $h_S h_T \cdot \text{rec}$ 

```

For a more efficient implementation, the discrete field quantities, e.g., $W_h = W(s, \mathcal{M}_X)$ or $\nabla \varphi_h$, are evaluated once on the entire mesh $\mathcal{M}_X$ and then accessed via the interpolation operator $\Pi: \mathbb{R}^{n_{\mathcal{M}_X}} \rightarrow \mathcal{C}(X)$, which performs bilinear interpolation of the grid values. All fundamental operations, such as the norm $|\cdot|$ or division $\cdot/\cdot$ on the discrete fields, have to be understood element wise.

C.4. Parallelization strategy. The formulation in eq. (41) naturally suggests a decomposition along the geometry-parameter axis S . Since each s_i -iteration is independent, we partition S into overlapping batches $\{S^{(r)}\}_{r=1}^P$ distributed across P processes. This overlap is a technical requirement to be able to apply finite-difference operators for some of the weight functions W , specifically those that include the partial derivative ∂_s . The batches $\{S^{(r)}\}_{r=1}^P$ are chosen to have (near-)equal size, ensuring balanced computational load across processes. Each process

then computes a partial reconstruction

$$(42) \quad \text{rec}^{(r)} = h_S h_T \sum_{s_i \in S^{(r)}} \sum_{t_j \in T} (\psi g)(s_i, t_j) \cdot v_{p, \gamma, W}(s_i, t_j),$$

using Algorithm 1, and the final reconstruction is obtained by reduction: $\text{rec} = \sum_{r=1}^P \text{rec}^{(r)}$. The overlapping regions are excluded from the local accumulation to avoid double-counting.

REFERENCES

- [1] G. BEYLKIN, *Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform*, J. Math. Phys., 26 (1985), pp. 99–108, <http://dx.doi.org/10.1063/1.526755>.
- [2] N. BLEISTEIN, *On the imaging of reflectors in the earth*, Geophysics, 52 (1987), pp. 931–942, <https://doi.org/10.1190/1.1442363>.
- [3] N. BLEISTEIN, J. K. COHEN, AND J. W. STOCKWELL, JR., *Mathematics of multidimensional seismic imaging, migration, and inversion*, vol. 13 of Interdisciplinary Applied Mathematics, Springer-Verlag, New York, 2001, <http://dx.doi.org/10.1007/978-1-4613-0001-4>. Geophysics and Planetary Sciences.
- [4] N. BLEISTEIN AND S. H. GRAY, *From the Hagedoorn imaging technique to Kirchhoff migration and inversion*, Geophysical Prospecting, 49 (2001), pp. 629–643, <https://doi.org/10.1046/j.1365-2478.2001.00290.x>.
- [5] J. K. COHEN AND N. BLEISTEIN, *Velocity inversion procedure for acoustic waves*, Geophysics, 44 (1979), pp. 1077–1085, <http://dx.doi.org/10.1190/1.1440996>.
- [6] R. COURANT AND D. HILBERT, *Methods of mathematical physics. Vol. II*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1989. Partial differential equations, Reprint of the 1962 original, A Wiley-Interscience Publication.
- [7] L. DEMANET AND R. J. HEWETT, *PySIT – Seismic Imaging Toolbox for Python*, Version v1.0, 2019, <http://pysit.org>.
- [8] R. FELEA, V. P. KRISHNAN, C. J. NOLAN, AND E. T. QUINTO, *Common midpoint versus common offset acquisition geometry in seismic imaging*, Inverse Probl. Imaging, 10 (2016), pp. 87–102, <https://doi.org/10.3934/ipi.2016.10.87>.
- [9] F. G. FRIEDLANDER, *Sound pulses*, Cambridge University Press, New York, 1958.
- [10] K. GANSTER, *A microlocal and visual comparison of 2D Kirchhoff migration formulas - software package*, 2024, <https://dx.doi.org/10.35097/YeHLCWQEafknUBDA>.
- [11] K. GANSTER, *The numerical scheme of approximate inverse for 2D linear seismic imaging*, PhD thesis, Karlsruher Institut für Technologie (KIT), 2025, <https://doi.org/10.5445/IR/1000182156>.
- [12] K. GANSTER, *A microlocal perspective on Kirchhoff migration - Software Package*, Version v1.0, Oct. 2026, <https://doi.org/10.35097/8vr8wwx4bu04062y>.
- [13] K. GANSTER, E. T. QUINTO, AND A. RIEDER, *A microlocal and visual comparison of 2D Kirchhoff migration formulas in seismic imaging*, Inverse Problems, 40 (2024), p. 115001, <https://doi.org/10.1088/1361-6420/ad797b>.
- [14] K. GANSTER AND A. RIEDER, *Approximate inversion of a class of generalized Radon transforms*, SIAM J. Imaging Sci., 16 (2023), pp. 842–866, <https://doi.org/10.1137/22M1512417>.
- [15] C. GRATHWOHL, *Seismic imaging with the elliptic Radon transform in 3D: analytical and numerical aspects*, PhD thesis, Karlsruhe Institute of Technology, 2020, <https://doi.org/10.5445/IR/1000105093>.
- [16] C. GRATHWOHL, P. KUNSTMANN, E. T. QUINTO, AND A. RIEDER, *Microlocal analysis of imaging operators for effective common offset seismic reconstruction*, Inverse Problems, 34 (2018), pp. 114001, 24, <https://doi.org/10.1088/1361-6420/aadc2a>.
- [17] C. GRATHWOHL, P. KUNSTMANN, E. T. QUINTO, AND A. RIEDER, *Imaging with the elliptic Radon transform in three dimensions from an analytical and numerical perspective*, SIAM J. Imaging Sci., 13 (2020), pp. 2250–2280, <https://doi.org/10.1137/20M1332657>.
- [18] V. GUILLEMIN, *On some results of Gel'fand in integral geometry*, in Pseudodifferential operators and applications (Notre Dame, Ind., 1984), vol. 43 of Proc. Sympos. Pure Math., Amer. Math. Soc., Providence, RI, 1985, pp. 149–155, <https://doi.org/10.1090/pspum/043/812288>.

- [19] V. GUILLEMIN AND S. STERNBERG, *Geometric asymptotics*, American Mathematical Society, Providence, R.I., 1977. Mathematical Surveys, No. 14.
- [20] J. G. HAGEDOORN, *A process of seismic reflection interpretation*, Geophysical Prospecting, 2 (1954), pp. 85–127, <https://doi.org/10.1111/j.1365-2478.1954.tb01281.x>.
- [21] P. HINTZ, *An introduction to microlocal analysis*, vol. 304 of Graduate Texts in Mathematics, Springer, Cham, Switzerland, 2025, <https://link.springer.com/book/9783031907050>.
- [22] P. C. KUNSTMANN, E. T. QUINTO, AND A. RIEDER, *Seismic imaging with generalized Radon transforms: stability of the Bolker condition*, Pure Appl. Math. Q., 19 (2023), pp. 1985–2036, <https://doi.org/10.4310/pamq.2023.v19.n4.a11>.
- [23] W. E. LORENSEN AND H. E. CLINE, *Marching cubes: A high resolution 3d surface construction algorithm*, SIGGRAPH Comput. Graph., 21 (1987), p. 163–169, <https://doi.org/10.1145/37401.37422>.
- [24] A. K. LOUIS, *Corrigendum: “Approximate inverse for linear and some nonlinear problems”* [Inverse Problems 11 (1995), no. 6, 1211–1223], Inverse Problems, 12 (1996), pp. 175–190, <http://dx.doi.org/10.1088/0266-5611/12/2/005>.
- [25] S. LUO AND J. QIAN, *Factored singularities and high-order Lax–Friedrichs sweeping schemes for point-source traveltimes and amplitudes*, J. Computat. Phys., 230 (2011), pp. 4742–4755, <https://doi.org/10.1016/j.jcp.2011.02.043>.
- [26] B. E. PETERSEN, *Introduction to the Fourier transform and pseudo-differential operators*, vol. 19 of Monographs and Studies in Mathematics, Pitman (Advanced Publishing Program), Boston, MA, 1983.
- [27] E. T. QUINTO, *Singularities of the X-ray transform and limited data tomography in $\mathbf{R}^2$ and $\mathbf{R}^3$* , SIAM J. Math. Anal., 24 (1993), pp. 1215–1225, <https://doi.org/10.1137/0524069>.
- [28] E. T. QUINTO AND A. RIEDER, *On microlocal artifacts arising in tomography modeled by a class of generalized Radon transforms*, Inverse Problems, (2026), <https://doi.org/10.1088/1361-6420/aeb624>.
- [29] E. T. QUINTO, A. RIEDER, AND T. SCHUSTER, *Local inversion of the sonar transform regularized by the approximate inverse*, Inverse Problems, 27 (2011), pp. 035006, 18, <http://dx.doi.org/10.1088/0266-5611/27/3/035006>.
- [30] A. RICHARDSON, *Deepwave*, Version v0.0.27, June 2026, <https://doi.org/10.5281/zenodo.3829886>.
- [31] J. SCHLEICHER, M. TYGEL, AND P. HUBRAL, *3-D true-amplitude finite-offset migration*, Geophysics, 58 (1993), pp. 1112–1126, <https://doi.org/10.1190/1.1443495>.
- [32] J. SCHLEICHER, M. TYGEL, AND P. HUBRAL, *Seismic true-amplitude imaging*, Society of Exploration Geophysicists, Tulsa, OK, USA, 2007, <https://doi.org/10.1190/1.9781560801672>.
- [33] T. SCHUSTER, *The method of approximate inverse: theory and applications*, vol. 1906 of Lecture Notes in Mathematics, Springer, Berlin, 2007, <http://dx.doi.org/10.1007/978-3-540-71227-5>.
- [34] W. W. SYMES, *Mathematics of reflection seismology*, tech. report, The Rice Inversion Project, Rice University, Houston, TX, USA, 1998, <http://www.trip.caam.rice.edu/downloads/preamble.pdf>.
- [35] R. VERSTEEG, *The Marmousi experience: Velocity model determination on a synthetic complex data set*, The Leading Edge, 13 (1994), pp. 927–936, <https://doi.org/10.1190/1.1437051>.