Laplace-Domain Crosstalk-Free Source-Encoded Elastic Full-Waveform Inversion Using Time-Domain Solvers

Zhaolun Liu*, Jürgen Hoffmann†, Etienne Bachmann‡, Congyue Cui‡, Frederik J. Simons‡, and Jeroen Tromp‡§

ABSTRACT

Crosstalk-free source-encoded elastic Full-Waveform Inversion (FWI) using time-domain solvers has demonstrated skill and efficiency at conducting seismic inversions involving multiple sources and receivers with limited computational resources. A drawback of common formulations of the procedure is that, by sweeping through the frequency domain randomly at a rate of one or a few sparsely sampled frequencies per shot, it is difficult to simultaneously incorporate time-selective data windows, as necessary for the targeting of arrivals or wave packets during the various stages of the inversion. Here, we solve this problem by using the Laplace transform of the data. Using complex-valued frequencies allows for damping the records with flexible decay rates and temporal offsets that target specific traveltimes. We present the theory of crosstalk-free source-encoded FWI in the Laplace domain, develop the details of its implementation, and illustrate the procedure with numerical examples relevant to exploration-scale scenarios.

INTRODUCTION

Full-waveform inversion (FWI) aims to use all the information in a seismogram to estimate subsurface material properties (Lailly, 1983; Tarantola, 1984, 1986). Over the last few decades, FWI has firmly established itself as an important inversion tool for both acoustic and elastic problems in active-source exploration seismology (Pratt, 1999; Pratt and Shipp, 1999; Plessix, 2006; Operto et al., 2013) as well as for regional and global problems using natural earthquakes as sources (Liu and Tromp, 2008; Liu and Gu, 2012; Tromp, 2020). Adjoint formulations of the inverse problem and spectral-element simulations have fast become an inseparable pairing (Tromp et al., 2008).

Full-Waveform Inversion

Most of the applications of FWI to real data acquired on land (e.g., Breinders and Pratt, 2007; Plessix et al., 2010; Lemaistre et al., 2019; Murphy et al., 2020) or in marine settings (Prieux et al., 2013a; Operto et al., 2015) have used the acoustic approximation to wavefield modeling. While reliable results can be obtained in that case, provided appropriate data preprocessing and inversion preconditioning are applied (Breinders and Pratt, 2007; Prieux et al., 2013a), elastic FWI is desirable for applications when the data include strong elastic effects. Acoustic FWI can lead to erroneous inversion results if applied to elastic data that are sensitive to strong velocity contrasts (Barnes and Charara, 2009; Mora and Wu, 2018; Pan et al., 2018; Pérez Solano and Plessix, 2023; Zhang et al., 2023). Elastic FWI is especially advantageous for the inversion of land seismic data (Pérez Solano and Plessix, 2019), marked by high-amplitude surface waves and elastic effects from near-surface heterogeneities. Elastic FWI also allows for the inversion of particle velocity data of multicomponent ocean-bottom seismometers that record compressional and shear waves directly, not just hydroacoustic pressure conversions. As shown by Cho et al. (2022), for example, the shallow-background shear velocity can be inverted from the horizontal components of motion recorded by ocean-bottom nodes. Multicomponent data help constrain the inversion process and mitigate ill-posedness (Prieux et al., 2013a,b; Pan et al., 2019).

A drawback of elastic FWI is that it remains computationally challenging to solve large-scale optimization systems with three-dimensional (3-D) simulations for hundreds or even thousands of sources. Frequency-domain solvers have proven to be efficient at computing the response of large numbers of

---

* Saudi Aramco, Dhahran, Saudi Arabia; Formerly Department of Geosciences, Princeton University, USA; zhaolun.liu@outlook.com.
† DNO, ASA, Norway.
‡ Department of Geosciences, Princeton University, USA.
§ Program in Applied and Computational Mathematics, Princeton University, USA.
sources (Pratt, 1999). These methods additionally allow for the selection of just a handful of contributing frequencies (Sigurgeirsson and Pratt, 2004), and for the straightforward and cheap incorporation of intrinsic attenuation in the form of complex-valued wave speeds (Marfurt, 1984). Despite these advantages, direct solvers scale poorly (Virieux and Operto, 2009), require vast amounts of storage, and thus are unsuitable for the large 3-D problems that we face today. Iterative frequency-domain solvers may not share these limitations (Operto et al., 2007; Plessix, 2009), but they do not have the benefit of source independence. In contrast, explicit time-domain solvers scale linearly with the number of sources owing to the absence of all-to-all communications (Komatitsch and Tromp, 2002; Komatitsch et al., 2002) and they do not command a large amount of memory. However, since the compute time for elastic FWI using time-domain solvers is proportional to the number of seismic sources (for anelastic kernels, see Komatitsch et al., 2016), distributing those over the processors of parallel computers claim large time blocks on huge computational platforms, in particular if the source parallelism is combined with a second level of parallelism by domain decomposition of the computational mesh.

Source Encoding

Source encoding remedies the run-away computational burden of wavefield modeling in seismic inversion and migration. It reduces the number of calculations by combining gathers from multiple shots or events into “encoded” “super-gathers” (Morton and Ober, 1998; Etgen, 2005; Zhang et al., 2005; Krebs et al., 2009; Tang and Biondi, 2009; Ben-Hadj-Ali et al., 2011; Schuster et al., 2011). Under the encoded multisource approach the computational cost of FWI or Reverse Time Migration (RTM) no longer grows linearly with the number of sources, allowing for the simultaneous consideration of several thousand sources.

The mathematical effect of source encoding is clear by inspection of the FWI gradient, the equivalent RTM imaging condition (Luo et al., 2009; Zhu et al., 2009; Luo et al., 2015), or the sensitivity kernel (Tromp et al., 2005). Disregarding weighting factors and other subtleties, these objects generally involve the correlation interaction between two scalar wavefields (Romero et al., 2000). Combining all contributions in a survey or experiment by summing the single-frequency wavefields individually over all distinct sources is computationally advantageous but creates cross-term artifacts. Crosstalk negatively affects the quality of the inversion gradient or migration image and may introduce inaccuracy in the model results. A source-encoding strategy amounts to applying an encoding operator prior to expansion over the sources in a manner such that the cross-terms become relatively unimportant.

For every one of the $N_s$ seismic sources or shot gathers, and for each of the $N_\omega$ frequencies under consideration, let $\tilde{u}_s(x, \omega_k)$ and $\tilde{u}_s^\dagger(x, \omega_k)$ be generic forward and adjoint wavefields (the star “$\dagger$” denotes complex conjugation) at position $x$ and angular frequency $\omega_k$, and $\alpha_s(\omega_k)$ a generic normalized encoding operator. Denoting the real part by $\Re$, the expanded imaging condition (Romero et al., 2000)

$$\tilde{K}(x) = \Re \sum_{k=1}^{N_s} \sum_{s=1}^{N_s} \alpha_s(\omega_k) \tilde{u}_s(x, \omega_k) \times \sum_{k'=1}^{N_s} \alpha_{s'}(\omega_k) \tilde{u}_{s'}^\dagger(x, \omega_k)$$

$$= \Re \sum_{k=1}^{N_s} \sum_{s=1}^{N_s} \tilde{u}_s(x, \omega_k) \alpha_{s'}(\omega_k) \tilde{u}_{s'}^\dagger(x, \omega_k)$$

$$+ \Re \sum_{k=1}^{N_s} \sum_{s=1}^{N_s} \sum_{s' \neq s} \tilde{u}_s(x, \omega_k) \alpha_{s'}(\omega_k) \tilde{u}_{s'}^\dagger(x, \omega_k)$$

suffers from crosstalk unless $\alpha_s(\omega_k)\alpha_{s'}(\omega_k) \rightarrow \delta_{ss'}\delta_{kk'}$. Encoding operators need to be chosen carefully to produce enough dissimilarity between all sources to mutually decorrelate them in order to reduce the effect of crosstalk (and random data) noise (Aghazade et al., 2022).

A number of encoding strategies have been proposed in the literature. Romero et al. (2000) suggest multiplying the shot gather by $e^{i\phi_s(\omega)}$ with a random phase $\phi_s(\omega)$. Note that the imaginary unit $i$ is to be distinguished from the $i$ used to denote a component index in what follows. Krebs et al. (2009) advocate randomizing polarity using $\alpha_s(\omega) = \pm 1$ drawn at random. Tang and Biondi (2009) and Schuster et al. (2011) propose shifting each shot with a random time offset. While they all suffer from residual crosstalk to some extent, these strategies furthermore require that the receiver spread be fixed for each shot, hence they are not applicable for, e.g., marine streamer data, with arrays that move with each shot. In contrast, a technique that is applicable to marine streamer acquisition is through plane-wave encoding, whereby shot gathers are linearly time-shifted with respect to source-receiver offset and summed to form a series of plane-wave gathers with different ray parameters (Etgen, 2005; Zhang et al., 2005; Vigh and Starr, 2008; Tang and Biondi, 2009). However, this presupposes a small shot interval to avoid aliasing, which may not be possible for realistic 3-D data sets, and a sufficiently large number of ray parameters.

Crosstalk-Free Source Encoding

Source-encoding strategies designed to be completely crosstalk-free have been developed in the last decade (Huang and Schuster, 2012, 2018; Krebs et al., 2013; Schuster and Huang, 2013; Zhang et al., 2018; Tromp and Bachmann, 2019). Both Huang and Schuster (2012) and Krebs et al. (2013) propose an encoding operator that is a narrowband filter, convolved in the time domain with each trace. Different shot gathers are allocated different filters from a nearly orthogonal set. In the context of least-squares RTM, Dai et al. (2013) propose assigning to each source a unique frequency and using an orthogonal encoding operator $\alpha_s(\omega_k) = e^{i\omega_k t}$. The encoded wavefields mix frequencies and sources $\{k, s\}$ to form “super-
gathers",
\[ u(x, t) = \Re \sum_{(k,s)=1}^{N_x \times N_s} e^{-i\omega_k t} \tilde{u}_s(x, \omega_k), \tag{3} \]
which can be computed in the time-domain by simultaneously activating multiple sources driven by specific frequencies (Nihei and Li, 2007). The encoded forward wavefield \( u(x, t) \) and its adjoint \( u^\dagger(x, t) \) are run past the “steady-state time” \( T \). The sensitivity kernel is obtained by zero-lag cross-correlation over \( \Delta \tau \), an interval over which the encoding operators are exactly orthogonal, which accomplishes the “deblending” or “decoding”:
\[ K(x) = \int_0^{\Delta \tau} u(x, T + \Delta \tau - t) u^\dagger(x, t + T) \, dt. \tag{4} \]
Both \( T \) and \( \Delta \tau \) require careful consideration, though authors often take \( T = \Delta \tau \) for convenience. Huang and Schuster (2018) apply this idea in acoustic FWI. Zhang et al. (2018, 2020) propose a hybrid method that calculates time-domain encoded wavefields but performs the kernel computation in the frequency domain.

Tromp and Bachmann (2019) provide a systematic and didactic overview of crosstalk-free source-encoded FWI methods. They make the distinction between the steady-state time \( T \) and the integration interval \( \Delta \tau \) required for decoding, which are not typically differentiated. They also formulate different misfits, e.g., for amplitude, phase, and double-difference measurements. Finally, they show that the relative reduction in computational cost of source-encoded to traditional FWI for a large number of sources is approximately given by the original duration of the simulated seismograms (e.g., about 5 s) times their bandwidth (e.g., 20 Hz), a relationship (e.g., a 100× reduction) that holds both for computation time and for the number of input/output (I/O) calls—per iteration (noting that source encoding might require more iterations). Their crosstalk-free source-encoded FWI method has been successfully applied at the global scale of earthquake tomography (Cui et al., 2023), and at the human scale of 3-D ultrasound tomography (Bachmann and Tromp, 2020).

The main drawback of source-encoded formulations arises from the difficulty of time-windowing modeled data when inverting one or a few sparsely sampled frequencies at one time: time- and frequency-limitation are intrinsically incompatible (Simons, 2010). Yet, time-windowing allows for the balanced selection of specific arrivals during the various stages of the inversion, which is often desirable in order to navigate the misfit surface towards a global optimum. In this paper, we propose the use of complex-valued frequencies through the Laplace transform (Shin et al., 2002; Bredners and Pratt, 2007; Brossier et al., 2009; Prieux et al., 2013a), which damps arrivals at a certain rate \( \gamma \) starting from a given traveltime \( t_0 \), to be defined below. We develop the theory of crosstalk-free source-encoded elastic FWI in the Laplace domain and illustrate it with numerical examples.

ELASTIC FWI IN THE LAPLACE DOMAIN

In this paper, we extend the theory of crosstalk-free source-encoded elastic full-waveform inversion. Our treatment largely follows the adjoint formulation laid out by Tromp et al. (2005) and Tromp and Bachmann (2019), but we enable damping of the seismogram beyond a certain time offset (both to target specific seismic phases and to ensure stability), by switching from the Fourier domain to the Laplace domain, involving complex frequencies.

For a generic time-domain function \( g(t) \), we write the one-sided Laplace transform at complex argument
\[ z = \gamma + i \omega, \tag{5} \]
where \( \gamma \) and \( \omega \) are real, as
\[ \bar{g}(z) = \mathcal{L}[g(t)](z) = \int_0^\infty g(t) e^{-zt} \, dt. \tag{6} \]

The objective function

Let us consider a source location \( x_s \) and a source-time function \( f(t) \) which drives a body-force point source acting in the \( \hat{n} \) direction with components
\[ f_j(x_s, t) = \hat{n}_j(x_s) f(t), \tag{7} \]
which gives rise to displacement components at a receiver located at \( x_r \) through convolution with the elastic Green’s function \( G_{ij}(x_r, x_s; t) \) in the form, whereby Einstein’s summation convention is implied,
\[ u_i(x_r, t) = \int_0^t G_{ij}(x_r, x_s; t - t') f_j(x_s, t') \, dt'. \tag{8} \]

With Laplace transforms of the synthetic, the source, and the Green’s function given by
\[ \bar{u}_i(x_r, z) = \mathcal{L}[u_i(x_r, t)], \tag{9} \]
\[ \bar{f}_j(x_s, z) = \mathcal{L}[f_j(x_s, t)], \tag{10} \]
\[ \bar{G}_{ij}(x_r, x_s; z) = \mathcal{L}[G_{ij}(x_r, x_s; t)]; \tag{11} \]
the Laplace convolution theorem transforms equation 8 to the equivalent expression, again summed over \( j \),
\[ \bar{u}_i(x_r, z) = \bar{G}_{ij}(x_r, x_s; z) \bar{f}_j(x_s, z). \tag{12} \]

The scaled Laplace coefficient of \( u_i(x_r, t) \), where the damping is given by \( \gamma \) and the shift is over the arbitrary constant \( t_0^r \), e.g., the first-arrival time of the elastic waves from source \( x_s \) to receiver \( x_r \), is given by
\[ \mathcal{L} \left[ e^{\gamma t_0^r} u_i(x_r, t) \right] (z) = e^{\gamma t_0^r} \bar{u}_i(x_r, z). \tag{13} \]

Because we will associate every source at \( x_s \) recorded by a receiver at \( x_r \) with a unique complex source frequency \( z_s \), we introduce the shorthand notation
\[ \bar{f}_j(x_s, z) \equiv \bar{f}_j(x_s, z_s), \tag{14} \]
\[ \bar{G}_{ij}(x_r, x_s) \equiv \bar{G}_{ij}(x_r, x_s; z_s), \tag{15} \]
\[ \bar{u}_i(x_r, z) \equiv e^{\gamma t_0^r} \bar{u}_i(x_r, z_s), \tag{16} \]
\[ \bar{d}_i(x_r, z) \equiv e^{\gamma t_0^r} \bar{d}_i(x_r, z_s). \tag{17} \]
where equation 17 applies to the Laplace transform of the observed data \( d_i(x_r, t) \).

The amplitude damping factor \( \gamma \) and the source-receiver dependent time shift \( t_0^{sr} \) (as might be picked from the data or calculated in an initial model) effectively allow for time windowing of the seismograms through the Laplace transform, ensuring Green’s function recovery through orthogonality beyond steady-state time, as will be clarified below. To ensure that the discrete Laplace transform of the observed data contains exactly the encoded complex frequencies \( z_s \), we may zero-pad or truncate the time series to be able to evaluate the Laplace transform

\[
\tilde{d}_{i}^{sr} = e^{\gamma t_0^{sr}} \int_{0}^{\Delta \tau} d_i(x_r, t)e^{-z_s t} dt,
\]

(18)

or truncate the time series to be able to evaluate the Laplace transform

\[
\int_{0}^{\Delta \tau} d_i(x_r, t)e^{-\gamma(t-t_0^{sr})}e^{-i\omega_s t} dt,
\]

(19)

over the integration interval \( \Delta \tau \), to be specified below.

The modeling residuals to be minimized are the differences between the scaled Laplace coefficients,

\[
\Delta \tilde{u}_{i}^{sr} = \tilde{u}_{i}^{sr} - \tilde{d}_{i}^{sr}.
\]

(20)

The overall misfit is defined as the sum of squares, with the “∗” denoting complex conjugation,

\[
\chi = \sum_{s=1}^{S} \sum_{r=1}^{R} \sum_{i=1}^{C} w_{sri} \Delta \tilde{u}_{i}^{sr} \Delta \bar{u}_{i}^{sr},
\]

(21)

accounting for \( s = 1, ..., S \) different frequencies or sources, \( r = 1, ..., R \) receivers, and \( i = 1, ..., C \) sensor components.

The weights \( w_{sri} \) are nonzero only for those receivers \( r \) that record a source (frequency) \( s \) on a component \( i \). Note the slight variation in notation compared to equations 2–3.

The objective of FWI is to adjust the Earth model, embodied by the Green’s function \( G \), in such a manner that synthetic predictions \( u \) made for observations \( d \) are ideally matched. When carried out in the least-squares sense, we require ways to minimize equation 21. See Appendix A for alternative misfit formulations.

**The misfit variation**

To perform the minimization of the misfit criterion \( \chi \) in equation 21 we begin by considering how it varies under Earth model perturbations. The variation of the misfit is

\[
\delta \chi = \partial_{\delta \rho} \sum_{s=1}^{S} \sum_{r=1}^{R} \sum_{i=1}^{C} w_{sri} \Delta \tilde{u}_{i}^{sr} \delta \tilde{u}_{i}^{sr}.
\]

(22)

Under the Born approximation (Nolet, 2008), in terms of the density and stiffness perturbations \( \delta \rho \) and \( \delta c_{ijklm} \), the time-domain perturbed displacement (Tromp et al., 2005)

\[
\delta u_t(x_r, t) = \int_{0}^{t} \int_{V} \left[ \delta \rho(x) G_{ij}(x_r, x; t - t') \partial_{z_s} u_j(x, t') \right. \\
+ \delta c_{ijklm}(x) \partial_{z_s} G_{ij}(x_r, x; t - t') \partial_{t} u_m(x, t') ] \ d^3x \ dt,
\]

(23)

whereby Einstein’s summation convention is now implied for the repeat indices \( j, k, l, m \), and from hereon out. Using the convolution identity and the derivative properties of the Laplace transform, the scaled Laplace coefficients of the perturbed wavefield are

\[
\delta \tilde{u}_{i}^{sr} = -e^{\gamma t_0^{sr}} \int_{V} \left[ \delta \rho(x) G_{ij}(x_r, x) z_s^2 \tilde{u}_j(x, z_s) \right.
\]

(24)

\[
+ \delta c_{ijklm}(x) \partial_{z_s} G_{ij}(x_r, x) \partial_{t} \tilde{u}_m(x, z_s) \right] \ d^3x,
\]

where initial conditions \( u_j(x, 0) = 0 \) and \( \partial_t u_j(x, 0) = 0 \) are used to reduce the Laplace transform of the second temporal derivative of the displacement.

Inserting the perturbed displacement \( \delta \tilde{u}_{i}^{sr} \) of equation 24 into equation 22 yields

\[
\delta \chi = -\partial_{\delta \rho} \int_{V} \left\{ \delta \rho(x) w_{sri} \Delta \tilde{u}_{i}^{sr} \tilde{G}_{ij}(x_r, x) z_s^2 \tilde{u}_j(x, z_s) \right. \\
+ \delta c_{ijklm}(x) w_{sri} \Delta \tilde{u}_{i}^{sr} \partial_{z_s} \tilde{G}_{ij}(x_r, x) \partial_{t} \tilde{u}_m(x, z_s) \right\} \ d^3x,
\]

(25)

with the summation convention now additionally applying over the indices \( s, r \) and \( i \), i.e., over all sources, receivers, and receiver components.

In the following, we focus on deriving the density kernel for simplicity. Elastic kernels can be derived in a similar way. In equation 25, guided by equations 12 and 7, we first substitute \( \tilde{u}_j(x, z_s) = \tilde{G}_{jm}(x, x_s) f^s n_m(x_s) \) and, denoting the Laplace coefficient of the source-time function as \( \tilde{f}^s \), we rewrite \( \delta \tilde{f} = -i \tilde{f}^s \) to obtain

\[
\delta \chi_\rho = \int_{V} \delta \rho(x) \Re \left\{ e^{\gamma t_0^{sr}} w_{sri} \Delta \tilde{u}_{i}^{sr} \tilde{G}_{ij}(x_r, x) \right. \\
\times z_s^2 \tilde{G}_{jm}(x, x_s) f^s n_m(x_s) \right\} \ d^3x' \]

(26)

\[
- \int_{V} \delta \rho(x) \Re \left\{ e^{\gamma t_0^{sr}} \Delta \tilde{u}_{i}^{sr} \tilde{G}_{ij}(x_r, x) z_s^2 \tilde{u}_j(x, z_s) \right\} \ d^3x',
\]

(27)

where we have defined a weighted residual as

\[
\Delta \tilde{u}_{i}^{sr} = -i w_{sri} f^s \tilde{u}_{i}^{sr},
\]

(28)

and redefined the Laplace coefficients of a new forward wavefield to be

\[
\tilde{u}_{i}^{sr}(x) = -i \tilde{G}_{jm}(x, x_s) n_m(x_s).
\]

(29)

Multiplying equation 27 by \( e^{\pm z_s (t-t')} \) and integrating over a suitably shifted time interval of interest allows us to exploit their orthogonality to rewrite equation 27 as

\[
\delta \chi_\rho = -\frac{2}{\Delta \tau} \int_{0}^{\Delta \tau} \int_{V} \left\{ \delta \rho(x) \right. \\
\times \left. \Re \left\{ e^{\gamma t_0^{sr}} \Delta \tilde{u}_{i}^{sr} \tilde{G}_{ij}(x_r, x) e^{-z_s(t+T)} \right. \right. \\
+ \Re \left( z_s^2 \tilde{u}_j'(x, t') e^{z_s(t+T)} \right) \right\} \ d^3x.
\]

(30)
Other than the damping parameter $\gamma$ and the trace-dependent temporal shift $t_0^\tau$, which provide opportunities for time-windowing, phase selectivity and numerical stability, equation 30 contains the length of the “decoding” integration interval $\Delta \tau$ which commences beyond the “steady-state” time shift $T$. These two parameters are briefly introduced in equation 4, but here as there, they await further identification and numerical specification.

### Interacting forward and adjoint wavefields

In the second factor (II) of equation 30 we recognize the second time derivative of an encoded forward wavefield component, summed over the sources, 

$$u_j(x, t + T) = \Re \left\{ \sum_{s=1}^{S} \hat{u}_j^*(x) e^{2\xi_s(t + T)} \right\},$$

(31)

into which we substitute equation 29, the definition of the temporal shift $\xi_s$, to obtain 

$$u_j(x, t + T) = \sum_{s=1}^{S} \int_{-\infty}^{T+t} G_{jm}(x, x_s, T + t - t') e^{2\xi_s(t + T)} \tilde{n}_m(x_s) e^{i\gamma t} \sin \omega_s \tau \, dt',$n_m(x_s) e^{i\gamma t} \sin \omega_s \tau \, dt',$$

(32)

that is, driven by sources $f_m(x_s, t)$ with source-time functions $f(t) = e^{i\gamma t} \sin(\omega_s t)$.

In the first factor (I) of equation 30, we recognize a wavefield, summed over all sources and receivers, which, using the definition of the Laplace transform, the reciprocity of the Green’s function, and a rearranging of the integration limits can be rewritten (see Appendix B) as

$$\phi_j(x, t + T) = \sum_{r=1}^{R} \int_{-\infty}^{T+\Delta \tau - t} G_{jr}(x, x_r, T + \Delta \tau - t - t')$$

$$\times \sum_{s=1}^{S} e^{\gamma \xi_s} \Re \left\{ \Delta u_s^r(t) e^{-\gamma(T + t)} \right\} \, dt',$$

(33)

where we now define an adjoint source that can be simplified (see Appendix B) to

$$f^j_k(x_r, t) = \Re \left\{ \sum_{s=1}^{S} e^{\gamma \xi_s} \Delta u_s^r(T + t) e^{i\omega_s(T + t)} \right\},$$

(34)

$$= \Re \left\{ \Delta u_s^r(T + t) e^{-\gamma(T + t)} \right\},$$

(35)

where we recognized the shifted inverse Fourier transform

$$\Delta u_s^r(t) = \sum_{i=1}^{\infty} e^{\gamma \xi_s} \Delta u_s^r(t) e^{i\omega_s t}. $$

(36)

We now introduce the adjoint wavefield

$$u_j^\dagger(x, t) = \sum_{r=1}^{R} \int_{0}^{T} G_{j}^{\dagger}(x, x_r, t - t') f_k^\dagger(x_r, t') \, dt',$$

(37)

from which we obtain the relationship

$$\phi_j(x, t + T) = u_j^\dagger(x, T + \Delta \tau - t).$$

(38)

Substituting equations 38 and 32 into equation 30 yields

$$\delta \chi_\rho = -\frac{2}{\Delta \tau} \int V \delta \rho(x)$$

$$\times \left( \int_0^{\Delta \tau} u_j^\dagger(x', T + \Delta \tau - t) \partial^2_x u_j(x, t + T) \, dt \right) \, d^3x.$$

As far as the density perturbation is concerned, equation 39 provides an accessible version for the misfit variation in equation 30. To gain further insight we now write it in the form that emphasizes the sensitivity kernel in the manner of Tromp et al. (2005).

### Fréchet derivative sensitivity kernels

Equation 39 is in the form

$$\delta \chi_\rho = \int V \delta \log \rho(x) K_\rho(x) \, d^3x,$$

(40)

which defines the Fréchet derivative with respect to the mass density as

$$K_\rho(x) = -\frac{2}{\Delta \tau} \int_0^{\Delta \tau} \rho(x) \, u_j^\dagger(x, T + \Delta \tau - t) \partial^2_x u_j(x, t + T) \, dt.$$ 

(41)

For the stiffness perturbations we ultimately have the similar expression

$$\delta \chi_c = \int V \delta c_{jklm}(x) K_{c_{jklm}}(x) \, d^3x,$$

(42)

containing, as in Tromp et al. (2005), the Fréchet kernel with respect to the elastic constants

$$K_{c_{jklm}}(x) = -\frac{2}{\Delta \tau} \int_0^{\Delta \tau} \epsilon^j_k(x, T + \Delta \tau - t)$$

$$\times \epsilon_{lm}(x, t + T) \, dt,$$

(43)

where no summation is implied, and $\epsilon_{lm} = \frac{1}{2} [\partial_l u_m + \partial_m u_l]$ and $\epsilon^j_k = \frac{1}{2} [\partial_j u^k_l + \partial_k u^j_l]$ denote the elements of the strain and the adjoint strain tensors, respectively.

We return to equation 25 for the total misfit variation, $\delta \chi = \delta \chi_\rho + \delta \chi_c$. In an isotropic medium with elastic tensor $c_{jklm} = (\kappa - 2\mu/3)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})$, equations 40 and 43 combine into the expression

$$\delta \chi = \int V \left[ \delta \log \rho(x') K_\rho(x') + \delta \log \kappa(x') K_\kappa(x') + \delta \log \mu(x') K_\mu(x') \right] \, d^3x', $$

(44)

where, in addition to the density kernel 41, we have the bulk and shear modulus kernels

$$K_\kappa(x) = -\frac{2}{\Delta \tau} \int_0^{\Delta \tau} \kappa(x) \partial_l u^j_k(x, T + \Delta \tau - t)$$

$$\times \partial_j u^l_k(x, t + T) \, dt,$$

(45)

$$K_\mu(x) = -\frac{2}{\Delta \tau} \int_0^{\Delta \tau} 2\mu(x) D_{lj}^k(x, T + \Delta \tau - t)$$

$$\times D_{ij}^k(x, t + T) \, dt,$$

(46)
whereby $D_{ij}(x, t) = \frac{1}{2}(\partial_t U_j + \partial_j U_i) - \frac{1}{3} \delta_k U_k \delta_{ij}$ and $D_{ij}^\dagger(x, t) = \frac{1}{2}(\partial_t U_j^\dagger + \partial_j U_i^\dagger) - \frac{1}{3} \delta_k U_k^\dagger \delta_{ij}$ are the strain deviators associated with the forward and adjoint wavefields.

The formalism developed so far, in particular equations 41 and 43 for the density and elastic kernels, require numerical evaluation. In particular, they require reconciliation with the philosophy espoused in the Introduction. Following equation 4, single-frequency wavefields calculations can be carried out, and subsequently recombined, to return gradients for the discussion of the steady-state time $T$ and the decoding interval $\Delta \tau$.

**COMPUTATIONAL CONSIDERATIONS**

While frequency-domain numerical modeling methods (Pratt, 1999) can be brought to bear on our problem, Nihei and Li (2007) propose an efficient time-domain method that assigns a single, unique, frequency to each shot and simulates all shots simultaneously without crosstalk. Their approach is key to our source-encoded full-waveform strategy (Zhang et al., 2018; Tromp and Bachmann, 2019), as we further motivate, and specify.

**Single-frequency time-domain modeling**

Let $G(t)$ denote a Green’s function for wave propagation from a certain source to a certain receiver, without specificity to simplify the notation. Its Fourier transform is

$$\tilde{G}(\omega) = \int_0^{\infty} G(t) e^{-i\omega t} \, dt,$$  

(47)

where $\omega$ denotes the continuous angular frequency. Over a finite time interval $\Delta \tau$, to be specified, the discrete angular frequency will be $\omega_k = 2\pi k/\Delta \tau$, for integer $k = 1, \ldots, K$. The relation between these time and frequency properties will be made explicit below.

The response to a single-frequency source-time function $f_k(t) = \cos(\omega_k t)$, through equation 47, can be equivalently expressed as

$$u_k(t) = \Re \left\{ \tilde{G}(\omega_k) e^{i\omega_k t} \right\},$$  

(48)

$$u_k(t) = \int_0^{\Delta \tau} G(t - t') \cos(\omega_k t') \, dt',$$  

(49)

$$= [G \ast f_k](t).$$  

(50)

The star “$\ast$” denotes convolution. Hence we can obtain $u_k(t)$ by solving the wave equation using a time-domain numerical solver driven by a monochromatic source-time function. The single-frequency $\tilde{G}(\omega_k)$ can be recovered via the inner product of equation 48 with respect to the $e^{\pm i\omega_k t}/\Delta \tau$ orthonormal basis for the interval $\Delta \tau$, shifted by the steady-state time $T$, to be defined. By virtue of orthogonality, this integration yields

$$\tilde{G}(\omega_k) \delta_{kk'} = \frac{2}{\Delta \tau} \int_T^{T+\Delta \tau} u_k(t) e^{-i\omega_k t} \, dt.$$  

(51)

The numerical evaluation of equation 51 may be performed using the Fast Fourier Transform (FFT), taking care of the factors $(2\Delta \tau/\Delta \tau) \exp(-i\omega_k T)$. In practice, and to increase numerical stability, we may also change the source-time function from $\cos(\omega_k t)$ to $\sin(\omega_k t)$, for which we multiply equation 51 by $i = \sqrt{-1}$.

The steady-state time $T$ is when the value of the Green’s function becomes vanishingly small, after which virtually no more seismic waves reach the recording station (Cui et al., 2023). Depending on the Earth model, the source-receiver configuration, and the numbers and types of the seismic sources, steady state may be difficult to achieve, a situation that can be remedied by the damping of the Green’s function as discussed further below.

Figure 1 shows a variety of example geometries. In Figure 1a there is precisely one source, at $x_1$, and only one receiver, at $x_r$. The single-frequency source-time function $f_1(t) = \cos(\omega_1 t)$, see Figure 2a, the Green’s function $G_1(t) = G(x_1, x_r; t)$, as shown in Figure 2b, and the seismic response, the convolution of $G_1(t)$ with $f_1(t)$, is $u(t) = u_1(t)$, as shown in Figure 2c. The Fourier terms of the Green’s function $G_1(\omega_k)$ at different frequencies $k$ are depicted by blue circles in Figure 2d, and $G_1(\omega_1)$, recovered via equation 51, is shown as a red diamond.

**Multiple-source encoding and decoding**

Figure 1b shows the situation where one receiver at $x_r$ is activated simultaneously by two sources at $x_1$ and $x_2$.
source-time functions for the two sources $f_1(t) = \cos(\omega_1 t)$ and $f_2(t) = \cos(\omega_2 t)$ are shown in Figures 2a and 3a, respectively. The Green’s functions to the receiver $x_r$ are $G_1(t) = G(x_r, x_1; t)$ and $G_2(t) = G(x_r, x_2; t)$, respectively, as shown in Figures 2b and 3b. The “supergather” $u(t)$ as shown in Figure 3c is the superposition of two terms of the form of equations 48–50,

$$u(t) = \sum_{k=1}^{2} u_k(t) = \sum_{k=1}^{2} G_k * f_k.$$  \hspace{1cm} (52)

Fourier coefficients $\tilde{G}_1(\omega_1)$ and $\tilde{G}_2(\omega_2)$ can be recovered without crosstalk according to equation 51. The recovery is illustrated numerically in Figure 3d and 3e.

Now suppose that there are $K$ frequencies within the frequency band $[\omega_{\text{min}}, \omega_{\text{max}}]$, with

$$\Delta \omega = (\omega_{\text{max}} - \omega_{\text{min}})/(K - 1),$$  \hspace{1cm} (53)

which, at last, defines the integration or decoding interval $\Delta \tau$ for the seismic signal as

$$\Delta \tau = 2\pi/\Delta \omega = \frac{2\pi(K - 1)}{\omega_{\text{max}} - \omega_{\text{min}}}.$$  \hspace{1cm} (54)

Our goal is to perform one “super” forward simulation combining all sources by effectively tagging each individual frequency. If there were as many physical sources as there are discrete frequencies, each source would be randomly assigned a monochromatic source time function with an angular frequency $\omega_k$ defined by

$$\omega_k = \omega_{\text{min}} + (k - 1)\Delta \omega, \quad k = 1, \ldots, K.$$  \hspace{1cm} (55)

In practical application, the set of frequencies determined by equation 55 is randomly distributed over the available sources at the start of every iteration. As $K$ is the number of frequencies rather than distinct sources, individual sources may be assigned more than one frequency. When the number of sources is very large, as with streamer data or large-scale nodal deployments, the frequency spacing can be reduced by increasing the interval $\Delta \tau$ without loss of orthogonality.

The duration of all observed seismic data may not be equal to $\Delta \tau$, in which case the frequency spacing of the observations may not equal that of the synthetics, $\Delta \omega$. In those cases, we may resort to truncation or zero-padding of the real data before Fourier transformation. If this unduly increases the computation time or storage requirements, we may first apply the discrete Fourier transform to the data directly, without zero-padding or truncating, and subsequently use Fourier interpolation to achieve similar results (Bachmann and Tromp, 2020). In the examples shown we use zero-padding or truncating, as required.

Figure 1c depicts the general case where multiple sources are recorded by multiple receivers, a common acquisition geometry in earthquake seismology or with marine streamer surveys where sources and receivers move for every shot, possibly combined with ocean-bottom node (OBN) deployments. Hence with simultaneous simulation, each receiver will record waves emanating from all sources. When there are two sources
and two receivers, the data recorded at $x_{r1}$ and $x_{r2}$ are

$$u_1(t) = \sum_{k=1}^{2} G_k \ast f_k,$$  \hspace{1cm} (56)$$

$$u_2(t) = \sum_{k=1}^{2} G_k \ast f_k,$$  \hspace{1cm} (57)$$

where $G_{k1}(t)$ and $G_{k2}(t)$ are the Green’s function from the source at $x_k$ to the receivers $x_{r1}$ and $x_{r2}$, respectively. Again, equation 51 shows how to integrate to $u_1(t)$ and $u_2(t)$ against $e^{-i\omega t}$ to obtain $G_{11}(\omega_1)$ and $G_{12}(\omega_1)$, respectively. Similarly, integrating against $e^{-i\omega t}$ will yield $G_{22}(\omega_2)$ and $G_{21}(\omega_2)$. Upon doing so we may choose to retain only $G_{11}(\omega_1)$ and $G_{22}(\omega_2)$, choosing to discard $G_{12}(\omega_1)$ and $G_{21}(\omega_2)$.

### Time-selective single-frequency modeling

In the methodology outlined above, the steady-state time $T$ should be reached for recovery of the single-frequency Green’s function. Reaching steady state may be challenging, to the detriment of source-encoded FWI. Damping by $\gamma > 0$ provides a solution. The Fourier transform of the damped Green’s function $G(t)$ is the Laplace transform

$$\tilde{G}(z) = \mathcal{L}[G(t)(z)] = \int_0^\infty G(t) e^{-\gamma t} e^{-izt} dt$$  \hspace{1cm} (58)$$

$$= \int_0^\infty G(t) e^{-zt} dt.$$

Next consider the response to a cosinusoidal source-time function with exponentially increasing amplitude, $f_k(t) = e^{\gamma t} \cos(\omega_k t) = \Re \{ e^{izt} \}$. The corresponding wavefield

$$u_k'(t) = \Re \{ \tilde{G}(z_k) e^{z_k t} \},$$

$$= \int_0^{2\pi} G(t - t') e^{-\gamma t'} \cos(\omega_k t') dt',$$

$$= [G \ast f_k'](t),$$

which again can be evaluated using a time-domain solver. In analogy with equation 51, the single complex-frequency Green’s function can now be recovered as

$$\tilde{G}(z_k) \delta_{kk'} = \frac{2}{2\pi} \int_T^{T + \Delta T} u_k'(t) e^{-z_{k'} t} dt.$$  \hspace{1cm} (64)$$

As discussed, to provide phase selectivity as well as numerical stability, usually, we will damp $G(t)$ by $e^{-\gamma(t-t_0^\gamma)}$, with $t_0^\gamma$ the arrival time for waves traveling from source $x_k$ to receiver $x_r$. The Fourier transform of the Green’s function damped in time starting from $t_0^\gamma$, assuming it is zero before that, is its simply scaled Laplace transform,

$$\mathcal{L} \left[ e^{\gamma t_0^\gamma} G(t) \right](z) = e^{z t_0^\gamma} \tilde{G}(z).$$  \hspace{1cm} (65)$$

The Laplace transform of the convolution between functions is the product of their Laplace transforms. The Laplace transform of the $n$th derivative of a function brings out the $n$th power of the argument, in particular,

$$\mathcal{L}[D_t^n u(t)](z) = z^n \tilde{u}(z) - z u(0) - D_t u(0).$$  \hspace{1cm} (66)$$

Figure 4 shows a worked example of single complex-frequency forward model. Figure 4a and 4b show source-time functions $f_1(t) = \cos(\omega_1 t)$ and $f_1'(t) = e^{\gamma t} \cos(\omega_1 t)$, Figure 4c and 4d the Green’s functions $G_1(t)$ and $G_1(t) e^{-\gamma(t-t_0^\gamma)}$, Figure 4e and 4f the wavefields, and Figure 4g and 4h the recovery of the single-frequency Fourier terms. Figure 4g shows the failure of decoding the Fourier coefficient $G_1(\omega_1)$ due to an inappropriate steady-state time $T$ for $G_1(t)$ in Figure 4c. For a damping factor $\gamma = 10 s^{-1}$, the damped Green’s function is shown in Figure 4d and the corresponding Fourier coefficients are shown as blue dashed stems in Figure 4h. Alternatively, we can convolve the source-time function $\cos(\omega_1 t) e^{\gamma t}$ in Figure 4b with $G_1(t)$ in Figure 4c to get the encoded data $u_1'(t)$ in Figure 4f. We can recover the Fourier coefficient at $\omega = \omega_1$ of the damped Green’s function using equation 64 (the red stem in Figure 4f).

Damping the Green’s function not only mutes late arrivals but also reduces the steady-state time. The time windowing
operation allows selection of specific arrivals during the stages of the inversion, as an effective means to mitigate the non-linearity of the inversion. For example, we may begin an inversion sequence in the Fourier domain, without any damping ($\gamma = 0$), in order to capture the large-amplitude surface waves and constrain the shallow velocity model, and later switch to the Laplace domain using damping ($\gamma > 0$) to focus on the early arrivals. In any case the damping factor $\gamma$ should be set appropriately so that the maximum value of the simulated wavefield does not exceed the valid range of the data type in which the calculations are carried out. A rule of thumb is $\gamma (T + \Delta \tau) < 22$ for 4-byte data (float32).

**Inversion workflow**

Our workflow is shown in Figure 5. We begin by transforming the observed data to the Laplace domain, choosing a damping parameter $\gamma$ and source-receiver-dependent time shift $t_{sr}^0$, obtaining scaled Laplace coefficients as in equations 17–19. Subsequently, we:

1. select random disjoint sets of scaled data Laplace coefficients for each source, $d_{sr}^r$;

2. activate all sources with source-time functions in the form of a (co)sine with amplitude exponentially increasing according with the (un-)damping parameter $\gamma$, as in equations 61–63;

3. carry out a source-encoded forward simulation until the wavefield reaches steady state after a time $T$, and obtain the encoded synthetic data that will be decoded owing to the orthogonality between the trigonometric (exponential) phase encoding terms;

4. encode the Laplace-domain data residuals 20 to obtain the adjoint source 34–36;

5. carry out one source-encoded adjoint simulation and calculate the gradients $K$ as in equations 41 and 45–46 through zero-lag cross-correlation between the steady-state encoded forward and adjoint wavefields over interval $\Delta \tau$, proportional to the inverse of the encoded frequency spacing $\Delta \omega$, as in equation 54.

**FREQUENCY ASSIGNMENT STRATEGIES**

Frequency-domain FWI is traditionally conducted by inverting single frequencies in succession from low to high, for all sources (Pratt, 1999; Sirgue and Pratt, 2004). In contrast, source-encoded FWI assigns to each source only one, or a few, frequencies at a time. A strategy needs to be designed to ensure that an adequate number of frequencies (indeed: all) are used for every source across the iterations. In the approach by Dai et al. (2013), each source is initially allocated a unique frequency within the band of interest. All source frequencies are increased by $\Delta \omega$ in subsequent iterations, wrapping around...
Figure 6: The Moving Frequency Band (MFB) method for source-encoding frequency assignment, a multiscale strategy that involves consecutive inversions of slightly overlapping frequency groups of fixed bandwidth. In this example, the shift $\delta \omega = \Delta \omega$, the frequency spacing. In each iteration, the frequencies assigned to the sources are randomly selected.

Figure 7: The multiscale strategy for source-encoded frequency assignment, a multiscale strategy that involves successive inversions of overlapping frequency groups whose lowest frequency is fixed. Each subsequent group adds one higher frequency. The frequencies assigned to the sources are randomly selected for each iteration.

Upon reaching the bandlimit, back to the minimum frequency. Only $(\omega_{\text{max}} - \omega_{\text{min}})/\Delta \omega$, iterations are needed to sample all frequencies for every source, hence there will be welcome sampling redundancy across the frequency spectrum.

In contrast to this type of systematic frequency sweep, we will adopt a random frequency assignment strategy, as proposed by Huang and Schuster (2018), Zhang et al. (2018) and Tromp and Bachmann (2019). Random frequency assignment is commonly combined with multiscale strategies in order to mitigate the nonlinearity inherent in FWI. We aim to utilize a wide frequency band to accommodate many sources without reducing the spacing $\Delta \omega$, which would increase simulation times. However, due to the nonlinearity of the inverse problem and under the threat of cycle-skipping, it may be necessary to use a narrower band in the early inversion stages. Hence the bandwidth will reflect a compromise between avoiding cycle-skips and simultaneously inverting as many sources and frequencies as possible. Ultimately, all are used.

Tromp and Bachmann (2019) propose the multiscale frequency selection strategy depicted in Figure 6. At the outset they choose a bandwidth and a frequency spacing $\Delta \omega$ that enables encoding all sources using equation 53. This initial interval is shifted up by $\delta \omega$ with each subsequent iteration, until the maximum frequency reaches the overall maximum desired. In our implementation, $\delta \omega$ could be a fraction of $\Delta \omega$, allowing for the bandwidth to stay constant for a set number of iterations. This approach is referred to as the “Moving Frequency Band” (MFB) method. Brossier et al. (2009) adapt the time-domain Bunks et al. (1995) approach to frequency selection as shown in Figure 7. They conduct successive inversions in widening frequency bands. The first iteration involves only the starting frequency interval, to which one frequency is added with every iteration. This approach is known as the “Bunks” method.

To further mitigate the nonlinearity of FWI, our Laplace-domain formulation allows for a subset of specific arrivals, e.g., early or reflected phases, to be used selectively, using the time offset $t_0$ and damping factors $\gamma$ by which to control the time windowing. The damping can be adjusted, e.g., to keep converted waves, free-surface multiples, or surface waves from adding to the nonlinearity of the inversion in the early iterations. For example, FWI of land data presents challenges due to increased nonlinearity caused by free-surface effects, including the propagation of surface waves, in the heterogeneous near-surface. With surface waves, multi-offset strategies may be necessary, as cycle-skipping accumulates nonlinearity with increasing offset (Liu and Huang, 2019; Borisov et al., 2020).

**NUMERICAL VALIDATION**

**Offshore Coupled Acoustic-Elastic FWI of Marmousi Streamer Data**

We demonstrate the applicability of our workflow to a synthetic streamer data set generated from the Marmousi model (Versteeg, 1994; Modrak and Tromp, 2016), overlain by a 450 m water layer. Both sources and receivers move with every shot. The shear-wave speed ($V_S$) model is derived from the compressional ($V_P$) model using a constant Poisson’s ratio of 0.25, $V_S = V_P / 1.732$. The density is assumed to be uniform at 1000 kg/m$^3$ and is not updated during the inversion. An absorbing boundary condition is applied around the model without adapting it to the free surface at the top of the water layer.

Figure 8a shows the compressional wave speed model, our inversion target. The initial $V_P$ and $V_S$ models for the inversion are formed by smoothing the true models with a 2-D Gaussian with vertical and horizontal standard deviations of 848.5 m, see Figure 8b. We used a lower starting frequency of 1.0 Hz, and employ the exponentiated phase cost function (Fu et al., 2018; Yuan et al., 2019) as defined in equation A-18. A first set of inversion tests is conducted without damping terms ($\gamma = 0$), that is, involving all the arrivals in the inversion.

The model dimensions are 9.2 km $\times$ 3.5 km. We simulate a 2-D towed-streamer seismic acquisition, consisting of 148 evenly distributed pressure sources excited 10 m below the water surface, with 61 m shot spacing. The streamer length is 8 km, and
includes 438 hydrophones with a spacing of 18 m at a depth of 10 m. The distance between the sources and the first recording channel is 30 m. The source-time function is a Ricker wavelet with a center frequency of 5 Hz, considered known. The maximum recording time is 7.5 s, the time step is 0.75 ms. The observed and synthetic data are computed with the 2-D spectral-element method SPECFEM2D (Komatitsch and Vilotte, 1998). The corresponding shot gather is depicted in Figure 9a.

Moving Frequency Band inversion, undamped

Using the MFB strategy, the first iteration follows Tromp and Bachmann (2019) in utilizing waves between 1–6 Hz. With each of the first 30 iterations, the frequency range shifts by 0.1 Hz. Iterations beyond 30 are performed in the range 4–9 Hz. Each supershot is 37.5 s long ($T + \Delta \tau$), while a non-encoded simulation only requires 7.5 s. For 150 sources and a bandwidth of 5 Hz, this resulted in a speed-up factor of 16 per iteration (see Tromp and Bachmann, 2019, their equation 58). The frequency assignment for each iteration and a frequency heatmap for all sources are shown in Figure 10a and 10c, respectively. Notice the uneven sampling in the low-frequency range. In Figure 11a solid lines represent the normalized model misfits, while the blue line in Figure 11b corresponds to the data phase misfit. The $V_P$ and $V_S$ models after 237 iterations are displayed in Figure 12a and 12c, respectively.

Bunks multiscale inversion, undamped

For the Bunks multiscale strategy we use the same time steps as with the MFB inversion. The first iteration employs waves in the 1–6 Hz range, increased by 0.1 Hz until iteration 30. The encoded source number increases until the 30th iteration. Iterations beyond 30 are performed between 1–9 Hz. A total of 240 frequencies are encoded simultaneously, ensuring that at least 90 sources are assigned two frequencies concurrently. Figure 10b and 10d show the frequency assignment per iteration and the heatmap for each frequency of all sources, respectively. Notice the more balanced sampling of the frequencies. The dashed lines in Figure 11a show the normalized model misfits for $V_P$ and $V_S$. The red line in Figure 11b shows the phase misfit curve. The Bunks approach uniformly covers the low frequencies, whereas the MFB strategy favors the higher frequencies. The preference for low frequencies pro-
provides a better update to the low-wavenumber components of the models, resulting in improved model updates. Figure 12b and 12d, respectively, show the $V_P$ and $V_S$ models that result, after 237 iterations.

**Bunks multiscale inversion, damped**

We repeat the experiment under the Bunks frequency assignment using the Laplace-domain approach, in three stages, starting with $\gamma = 2.0 \text{ s}^{-1}$, then $\gamma = 0.8 \text{ s}^{-1}$, and finally without damping, $\gamma = 0$, which is equivalent to using the Fourier transform. The damped common-shot gathers are shown in Figure 9b and 9c. The inversion parameters of the three stages are presented in Table 1. Note that damping slows the rate of convergence, offsetting some of the gains in speed made by source-encoding, and thus needs to be carefully considered.

For the damping term $\gamma = 2.0 \text{ s}^{-1}$, the steady-state time $T$ and the decoding interval $\Delta \tau$ are maintained at the common value, 5.25 s, and only $V_P$ is updated because $V_S$ is not sensitive to early arrivals. For $\gamma = 0.8 \text{ s}^{-1}$, $T$ and $\Delta \tau$ are set to 7.5 s. As the number of encoded sources is always smaller than the total number of sources available, only a "mini-batch"
Table 1: Inversion parameters for the Laplace-domain source-encoded FWI of streamer data in the Marmousi model.

<table>
<thead>
<tr>
<th>stage</th>
<th>$\gamma$ ($s^{-1}$)</th>
<th>band (Hz)</th>
<th>$\delta\omega$</th>
<th>$\omega_{\text{max-shift}}$</th>
<th>$T$ (s)</th>
<th>$\Delta\tau$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>0.08</td>
<td>5.25</td>
<td>5.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>1–3</td>
<td>0.08</td>
<td>6</td>
<td>7.50</td>
<td>7.50</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.50</td>
<td>7.50</td>
<td>30.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 13: Model misfit evolution for the Laplace-domain FWI of the streamer data for the three-stage inversion tabulated in Table 1.

Figure 14: Result after 1600 iterations using the Bunks approach of the three-stage Laplace-domain FWI of Marmousi streamer data: inverted (a) $V_P$ and (b) $V_S$ models.

Offshore Couple Acoustic-Elastic FWI of Marmousi Multi-Component OBN Data

In this section, we will examine the Laplace-domain source-encoded acoustic-elastic FWI for the inversion of a multi-component ocean-bottom node (OBN) dataset. The coupled acoustic-elastic spectral-element implementation can accurately simulate the full wavefield at and near the seafloor. It correctly handles fluid-solid boundary conditions (Komatitsch et al., 2000), which has significant advantages for imaging and inversion applications that use amplitude information for model building.

We will investigate how the data damping can affect the inversion process. The true and initial Marmousi models are again those already depicted in Figure 8. The source is positioned 10 m below the sea surface and a total of 150 sources are evenly distributed with a spacing of 60 m. At the seafloor, 46 OBN receivers are placed with a spacing of 200 m and two components per node are recorded. Exploiting reciprocity, as there are twice more physical sources than OBN receivers, it is more efficient to perform two simulations per vertical and horizontal geophone (treated as vertical and horizontal forces) to build the gradient. Therefore, a total of 92 virtual sources will be used. The recording time spans 11.25 s with a time step of 0.75 ms. The simulated common-receiver gather for the vertical and horizontal components is shown in Figure 15a and 15b, respectively. Figures 15c–f show the damped wavefields.
Figure 15: Seismograms for the vertical (left column) and horizontal (right column) data computed in the Marmousi model with different damping factors: (a–b) zero damping, $\gamma = 0 \, \text{s}^{-1}$, (c–d) $\gamma = 1 \, \text{s}^{-1}$, (e–f) $\gamma = 2 \, \text{s}^{-1}$. Time damping is applied from the first arrival onward to preserve long-offset information.

We first apply the Bunks multiscale strategy without damping, $\gamma = 0$. The first iteration utilizes waves in the 1–3 Hz range with a steady-state time $T = 7.5 \, \text{s}$ and an integration interval $\Delta \tau = 30 \, \text{s}$. In the initial iteration, 60 of the virtual sources are randomly chosen for inversion. For each subsequent iteration, the maximum frequency is increased by 0.1 Hz until reaching iteration 60. After that, iterations are performed between 1–9 Hz. A total of 240 frequencies are encoded simultaneously, allowing some sources to be assigned two or three frequencies at the same time. The normalized model misfits are shown in Figure 16a. The $V_P$ and $V_S$ models resulting after 700 iterations are displayed in Figure 17a and 17c, respectively.

We repeated the experiment with two damping terms, $\gamma = 2.0 \, \text{s}^{-1}$ and $1.0 \, \text{s}^{-1}$. The inversion parameters are unchanged from Table 1, except $\gamma = 1 \, \text{s}^{-1}$ in the second stage. The damped vertical and horizontal gathers are shown in Figure 15. The model misfits are shown in Figure 16b. The final $V_P$ and $V_S$ models, obtained using the damped Laplace-domain source-encoded FWI method, are found in Figures 17b and 17d.

In Figure 16a, using (undamped) Fourier-domain source-encoding of OBN data, we notice that the $V_S$ model receives more updates than $V_P$. In contrast, with the inversion of streamer data in Figure 11a, we observed more updates to $V_P$. This behavior may indicate that the OBN data contain more $S$-wave information beneficial to $V_S$ modeling. However, in real-world situations, we may suffer from bad initial $V_S$ models leading the inversion to a local minimum. Thus, a Laplace-domain method is necessary to fit the early arrivals, mainly inverting $V_P$, and then gradually updating the $V_S$ model as shown in Figure 16.

We perform these tests on a cluster, Della, equipped with an 80 GB NVIDIA A100 GPU and forty-eight 2.8 GHz 1000 GB Intel Ice Lake CPUs with 1000 GB of memory. Compute times for stages 1, 2, and 3 are 59.6 s, 64.46 s, and 105.6 s, respectively, per iteration.

Onshore Elastic FWI of Foothills Data

In this section, we will demonstrate the effectiveness of the Laplace-domain source-encoded elastic FWI method for land data. FWI of land data poses challenges due to increased non-linearity introduced by free-surface effects such as the propagation of surface waves in the heterogeneous near-surface. Additionally, the presence of short wavelengths in the shear wavefield necessitates an accurate $V_S$ starting model, particularly...
To illustrate the efficacy of Laplace-domain source-encoded elastic FWI, we consider a 2-D onshore Foothills model (Brenders et al., 2008) measuring 5.5 km × 15 km. The model incorporates various challenging geological features, including rough topography, alluvial surface deposits, and complex structures resulting from compressive fold-and-thrust tectonics associated with mountain building. The $V_P$ model shown in Figure 18a is modified from Figure 2a of Brenders et al. (2008). A corresponding $V_S$ model is constructed using a constant Poisson ratio, i.e., $V_S = V_P/1.732$. A uniform density of 2600 kg/m$^3$ is assumed to be known during the inversion. A free-surface is imposed at the top of the model.

We simulate an onshore survey with 300 explosive sources spaced every 50 m, 25 m below the surface. Recordings are made on the surface with 601 vertical geophones spaced every 25 m. Observed and synthetic data are computed with an iden-
Figure 20: Seismograms for the vertical data computed in the Foothills model with different damping factors: (a) $\gamma = 0$ s$^{-1}$, (b) $\gamma = 3$ s$^{-1}$, (c) $\gamma = 2.35$ s$^{-1}$, (d) $\gamma = 1$ s$^{-1}$, and (e) $\gamma = 0.8$ s$^{-1}$. We mark the first-break times in (a) and show two offset ranges, $t_{0r}^{br} \leq 1$ s and $t_{0r}^{br} \leq 2$ s, respectively.

CONCLUSIONS

To reduce the computational time of elastic full-waveform inversion (FWI) using time-domain solvers, we develop a new...
Table 2: Inversion parameters for the (un)damped source-encoded FWI of land data in the Foothills model.

<table>
<thead>
<tr>
<th>stage</th>
<th>$\gamma$ (s$^{-1}$)</th>
<th>band (Hz)</th>
<th>$\delta \omega$</th>
<th>$\omega_{\text{max-shift}}$</th>
<th>$T$ (s)</th>
<th>$\Delta \tau$ (s)</th>
<th>max $t_{0r}^n$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier</td>
<td>1</td>
<td>1.0</td>
<td>1–3</td>
<td>0.1</td>
<td>5</td>
<td>8.0</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>1–3</td>
<td>0.0063</td>
<td>5</td>
<td>3.0</td>
<td>3.6</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>2.35</td>
<td>1–3</td>
<td>0.05</td>
<td>5</td>
<td>4.0</td>
<td>4.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Laplace</td>
<td>6</td>
<td>1.7</td>
<td>1–3</td>
<td>0.1</td>
<td>7</td>
<td>6.0</td>
<td>7.0</td>
</tr>
<tr>
<td>7</td>
<td>0.8</td>
<td>1–3</td>
<td>0.1</td>
<td>8</td>
<td>7.5</td>
<td>18.5</td>
<td>99</td>
</tr>
<tr>
<td>Fourier</td>
<td>8</td>
<td>0.0</td>
<td>1–6</td>
<td>0.1</td>
<td>5</td>
<td>8.0</td>
<td>60</td>
</tr>
</tbody>
</table>

Figure 21: Inverted (a-b) $V_P$ and (c-d) $V_S$ models by (a,c) Fourier (undamped) and (b,d) damped Laplace-domain source-encoded FWI of Foothills land data.

version of source encoding that assigns to each source at random for each iteration a unique complex frequency, and that includes a damping factor to attenuate late arrivals. The misfit criterion is the sum of squared errors in the scaled Laplace coefficients between observed and synthetic data. The source-time function takes the form of a weighted cosine or sine with exponentially increasing amplitude.

We simultaneously activate all the sources and carry out one source-encoded forward simulation followed by one source-encoded adjoint simulation, using a time-domain solver. The encoded forward and adjoint wavefields are run past their steady state. The gradient is calculated through zero-lag cross-correlation between the steady-state encoded forward and adjoint wavefields over a decoding time interval proportional to the inverse of the encoded frequency spacing. Owing to the orthogonality between the trigonometric terms of the encoding operator, no crosstalk is introduced during gradient calculation, and there are no requirements or limitations on acquisition geometry. By tuning the damping factor, we can time-window the data even when only one or a few sparse frequencies are being sampled. Time-windowing allows for the selection of specific arrivals during the various stages of the inversion.

The new version of source encoding can reduce the computational cost of elastic FWI with time-domain solvers. Using our method, a practicing geoscientist can implement elastic FWI on a small workstation. All shot gathers are "encoded" into one "supergather", instead of needing to distribute the seismic sources over the processors of multiple computers in parallel systems that require high-performance computational platforms.

ACKNOWLEDGEMENTS

We thank the three anonymous reviewers, the Associate Editor, Scott Keating, and the Assistant Editor, Yang Liu, for their constructive comments and productive advice. We thank the Princeton Institute for Computational Science and Engineering (PICSciE) for computational facilities and support.

APPENDIX A

ALTERNATIVE MISFITS

We consider three formulations in addition to the misfit criterion in equation 21.

**Phase measurement**

We define

$$s^a(x_r, t_{0r}^n) = A^a(x_r) \exp(\imath \theta^a(x_r)),$$

$$d^a(x_r, t_{0r}^n) = A_{\text{obs}}^a(x_r) \exp(\imath \theta_{\text{obs}}^a(x_r)),$$
where
\[
\theta^s(x_r) = \arctan \left[ \frac{\mathcal{I}\{s^s(x_r, t_{0r}^s)\}}{\mathcal{R}\{s^s(x_r, t_{0r}^s)\}} \right], \tag{A-3}
\]
and
\[
A^s(x_r) = |s^s(x_r, t_{0r}^s)| = \sqrt{[\mathcal{R}\{s^s(x_r, t_{0r}^s)\}]^2 + [\mathcal{I}\{s^s(x_r, t_{0r}^s)\}]^2}, \tag{A-5}
\]
\[
A_{obs}^s(x_r) = |d^s(x_r, t_{0r}^s)| = \sqrt{[\mathcal{R}\{d^s(x_r, t_{0r}^s)\}]^2 + [\mathcal{I}\{d^s(x_r, t_{0r}^s)\}]^2}. \tag{A-6}
\]

The phase misfit function is defined as
\[
\chi_\theta = \frac{1}{2} \sum_{s=1}^{S} \sum_{r=1}^{R_s} (\Delta \theta_r^s)^2, \tag{A-7}
\]
where \(\Delta \theta_r^s\) is the differential phase measurement,
\[
\Delta \theta_r^s = \arctan \left[ \frac{\mathcal{I}\{s^s(x_r, t_{0r}^s)\}}{\mathcal{R}\{s^s(x_r, t_{0r}^s)\}} \right] - \theta_{obs}^s(x_r), \tag{A-9}
\]
where the last equality holds modulo 2\(\pi\). This phase misfit function has variation,
\[
\delta \chi_\theta = \mathcal{R} \sum_{s=1}^{S} \sum_{r=1}^{R_s} \Delta \theta_r^s \delta \theta^s(x_r), \tag{A-10}
\]
\[
\mathcal{R} \sum_{s=1}^{S} \sum_{r=1}^{R_s} \Delta \theta_r^s [A^s(x_r)]^{-2} \times [-is^{ss}(x_r, t_{0r}^s)] \delta s^s(x_r, t_{0r}^s), \tag{A-11}
\]
where we can define the weighted data residual as
\[
\Delta s^s(x_r, t_{0r}^s) = \Delta \theta_r^s [A^s(x_r)]^{-2} s^{ss}(x_r, t_{0r}^s) f^s, \tag{A-12}
\]
so that
\[
\Delta s^s(x_r, t_{0r}^s) = \Delta \theta_r^s [A^s(x_r)]^{-2} s^{ss}(x_r, t_{0r}^s) f^s, \tag{A-14}
\]
so we can get the adjoint
\[
\hat{f}^s(x_r, t) = f^s(x_r, T + \Delta t - t), \tag{A-15}
\]
where we can define
\[
\Delta s^s_t(x_r, t_{0r}^s) = e^{it_{0r}^s} \Delta s_{s}^s(x_r, t_{0r}^s). \tag{A-17}
\]
Exponentiated Phase

We also consider the exponentiated phase misfit function (Yuan et al., 2019),

\[ \chi_\theta = \frac{1}{2} \sum_{s=1}^{S} \sum_{r=1}^{R_s} \left| \exp[i\theta^s(x_r)] - \exp[i\theta^s_{obs}(x_r)] \right|^2 \]  
\[ = 2 \sum_{s=1}^{S} \sum_{r=1}^{R_s} \sin^2 \left( \frac{1}{2} \Delta \theta^s_r \right), \]  
(19)

and its perturbation

\[ \delta \chi_\theta = \sum_{s=1}^{S} \sum_{r=1}^{R_s} \sin(\Delta \theta^s_r) \delta \theta^s(x_r), \]  
(20)

\[ = \mathcal{R} \sum_{s=1}^{S} \sum_{r=1}^{R_s} \sin(\Delta \theta^s_r) [A^s(x_r)]^{-2} \times [-i \Delta s^s(x_r, t_0^r)] \delta s^s(x_r, t_0^r). \]  
(21)

Now define

\[ \Delta s^s(x_r, t_0^r) = \sin(\Delta \theta^s_r) [A^s(x_r)]^{-2} [i \Delta s^s(x_r, t_0^r)] (if^s), \]  
\[ = \sin(\Delta \theta^s_r) [A^s(x_r)]^{-2} \Delta s^s(x_r, t_0^r) f^s, \]  
(22)

so that

\[ \Delta s^s(x_r, t_0^r) = \sin(\Delta \theta^s_r) [A^s(x_r)]^{-2} \Delta s^s(x_r, t_0^r) f^s. \]  
(23)

We can then get the adjoint,

\[ \tilde{f}^i(x_r, t) = f_i^i(x_r, T + \Delta t - t), \]  
\[ = \mathcal{R} \left\{ \sum_{s=1}^{S} \Delta \hat{s}^s(x_r, t_0^r) e^{i\omega_s(t + T)} \right\} e^{-i \gamma(T + t)}, \]  
(25)

where

\[ \delta \hat{s}^s(x_r, t_0^r) = e^{i\gamma t_0^r} \Delta \hat{s}^s(x_r, t_0^r). \]  
(26)

Double-Difference Phase

We next consider the misfit function (Yuan et al., 2016)

\[ \chi_\theta^{DD} = \frac{1}{2} \sum_{s=1}^{S} \sum_{r=1}^{R_s} \sum_{r' > r} \left[ \Delta \theta^s_{rr'} \right]^2, \]  
(27)

where \( w_{rr'} \) is a suitably chosen weighting function, and \( \Delta \theta^s_{rr'} \) is the “double-difference” phase measurement,

\[ \Delta \theta^s_{rr'} = \Delta \theta^s_r - \Delta \theta^s_{r'}, \]  
(28)

so we can get the adjoint,

\[ \tilde{f}^i(x_r, t) = f_i^i(x_r, T + \Delta t - t) \]  
\[ = \mathcal{R} \left\{ \sum_{s=1}^{S} \delta \hat{s}^s(x_r, t_0^r) e^{i\omega_s(t + T)} \right\} e^{-i \gamma(T + t)}, \]  
(39)

where we can define

\[ \delta \hat{s}^s(x_r, t_0^r) = e^{i\gamma t_0^r} \Delta \hat{s}^s(x_r, t_0^r). \]  
(40)
APPENDIX B

SIMULTANEOUS-SOURCE FORWARD AND ADJOINT WAVEFIELD

Forward Wavefield and Source

To reduce the second factor (II) of equation 30, we derive, via equation 31

\[ u_j(x, t + T) = \mathcal{R} \left\{ \bar{u}_j^n(x) e^{z_j(t+T)} \right\}, \]  

where the source term is defined as

\[ \phi_j(x, t + T) = \mathcal{R} \left\{ e^{\gamma t_0} \bar{\Delta u}_i^{s*} e^{-z_i(t+T)} G_{ij}(x_r, x) \right\}, \]  

for \( t \in [0, \delta t] \).

\[ \phi_j(x, t + T) = \mathcal{R} \left\{ e^{\gamma t_0} \bar{\Delta u}_i^{s*} e^{-z_i(T+t)} \right\} \times \int_0^\infty G_{ji}(x, x_r, t') e^{-z_i t'} dt', \]  

where the adjoint source term is defined as

\[ f_i^j(x_r, T + \Delta \tau - t') = \sum_{s=1}^S e^{\gamma t_0} \mathcal{R} \left\{ \bar{\Delta u}_i^{s*} e^{-z_i(2T+\Delta \tau-t')} \right\} \text{adjoint source}, \]  

\[ \phi_j(x, t + T) = \sum_{r=1}^R \int_{t-\Delta t}^{t+\Delta t} G_{ji}(x, x_r, T + \Delta \tau - t') \times \sum_{s=1}^S e^{\gamma t_0} \mathcal{R} \left\{ \bar{\Delta u}_i^{s*} e^{-z_i(2T+\Delta \tau-t')} \right\} dt', \]  

Adjoint Wavefield and Adjoint Source

To reduce the first factor (I) of equation 30, we define a wavefield, summed over all sources and receivers, using the Laplace transform definition and the reciprocity of the Green’s function, and for \( t \in [0, \delta t] \),

\[ \phi_j(x, t + T) = \mathcal{R} \left\{ e^{\gamma t_0} \bar{\Delta u}_i^{s*} e^{-z_i(T+t)} \right\} \times \int_0^\infty G_{ji}(x, x_r, t') e^{-z_i t'} dt', \]  

Adjoint Wavefield and Source

To reduce the first factor (I) of equation 30, we define a wavefield, summed over all sources and receivers, using the Laplace transform definition and the reciprocity of the Green’s function, and for \( t \in [0, \delta t] \),

\[ \phi_j(x, t + T) = \mathcal{R} \left\{ e^{\gamma t_0} \bar{\Delta u}_i^{s*} e^{-z_i(T+t)} \right\} \times \int_0^\infty G_{ji}(x, x_r, t') e^{-z_i t'} dt', \]  

where the adjoint source term is defined as

\[ f_i^j(x_r, T + \Delta \tau - t') = \sum_{s=1}^S e^{\gamma t_0} \mathcal{R} \left\{ \bar{\Delta u}_i^{s*} e^{-z_i(2T+\Delta \tau-t')} \right\}, \]  

"REFERENCES"


574, doi: 10.1190/1.3124932.
Tang, Y., and B. Biondi, 2009, Least-squares migra-
Zhang, Z., Z. Wu, Z. Wei, J. Mei, R. Huang, and P. Wang, 2023, Enhancing salt model resolution and subsalt imag-