

Source encoding for adjoint tomography

Jeroen Tromp^{1,2} and Etienne Bachmann¹

¹*Department of Geosciences, Princeton University, Guyot Hall, Princeton, NJ 08544, USA. E-mail: jtromp@princeton.edu*

²*Program in Applied & Computational Mathematics, Princeton University, Guyot Hall, Princeton, NJ 08544, USA*

Accepted 2019 June 6. Received 2019 June 4; in original form 2019 May 15

SUMMARY

We use a version of source encoding to facilitate the calculation of the gradient of a misfit function independent of the number of sources or receivers. The crosstalk-free method requires only two numerical simulations per iteration, namely, one source-encoded forward simulation and one source-encoded adjoint simulation. Importantly for practical applications, each source does not need to be recorded by all receivers. The method is implemented in two complementary ways. In the first approach the encoded forward and adjoint wavefields are run until they reach steady state, at which point they are ‘decoded’ to obtain their stationary parts. These stationary parts are combined to calculate the misfit gradient by summing their respective contributions. In the second approach the steady-state encoded forward and adjoint wavefields are convolved over a time period proportional to the inverse of the encoded frequency spacing. Using this strategy, the encoded forward and adjoint wavefields do not need to be decoded, nor is there a need to calculate or store intermediary stationary contributions to the gradient. We consider a wide variety of source-encoded misfit functions, including waveform differences, phase and amplitude measurements, ‘double-difference’ phase and amplitude measurements, cross-correlation traveltimes measurements, and a generic adjoint tomography misfit function. When measurements in specific time windows are involved in the construction of the source-encoded misfit function, as in adjoint tomography, the computational cost scales linearly with the number of seismic sources, because the necessary synthetic seismograms must be computed individually. In contrast, when using ‘super measurements’ based on source-encoded Fourier coefficients of entire observed and simulated seismograms, as in pure full waveform inversion, one iteration requires just two numerical simulations, independent of the number of sources and receivers. We illustrate the method based on examples from both earthquake and exploration seismology, highlighting inversion options and strategies involving frequency- and time-domain encoding, decoding (with more than 16 000 frequencies), encoded frequency randomization, encoding multiple frequencies per source, effects of noise, a variable number of receivers per event, various measurements and related misfit functions and attenuation.

Key words:;;

1 INTRODUCTION

Full waveform inversion (FWI) is a physics-based technique for solving seismic inverse problems (Tarantola 1984b, 1986). Thanks to rapid advances in high-performance computing, in recent years FWI has become an important 3-D inversion tool (e.g. Pratt *et al.* 1998; Pratt 1999; Pratt & Shipp 1999; Plessix 2006; Virieux & Operto 2009; Operto *et al.* 2014). In earthquake seismology, a variant of FWI based on the adjoint-state method, sometimes called ‘adjoint tomography’, has enabled 3-D inversions on regional (Tape *et al.* 2009, 2010; Fichtner *et al.* 2009; Luo *et al.* 2009; Zhu *et al.* 2012; Luo *et al.* 2013; Zhu & Tromp 2013; Chen *et al.* 2015) and global (Bozdağ *et al.* 2016; Lei *et al.* 2019) scales.

The computational cost of FWI and adjoint tomography is considerable, involving 3-D simulations for hundreds or even thousands of events (shots or earthquakes) recorded by thousands of receivers. Direct frequency-domain solvers have the significant benefit that they are independent of the number of sources (e.g. Pratt 1999). Unfortunately, these solvers rely on LU factorization, a numerical technique well known to poorly scale and requiring a substantial amount of memory for 3-D problems. Iterative frequency-domain solvers do not have this limitation, but they do not have the benefit of source independence. In contrast, explicit time-domain solvers scale linearly, thanks to the

absence of all-to-all communications (e.g. Komatitsch & Tromp 1999; Komatitsch *et al.* 2002). Inversion and assembly of a global system matrix can be avoided, thereby significantly reducing the memory requirements.

From the perspective of the seismic inverse problem, time-domain solvers have two main drawbacks, namely, time step limitations imposed by the CFL condition and a compute time proportional to the number of seismic sources. The first issue is mitigated by requirements of accuracy for the calculation of Fréchet derivatives. The second issue is an active area of research.

In exploration seismology, the linear dependence of the cost of the inverse problem on the number of sources has been dealt with based on an approach referred to as ‘source encoding’ (Krebs *et al.* 2009; Ben-Hadj-Ali *et al.* 2009; Choi & Alkhalifah 2011; Schuster *et al.* 2011; Schiemenz & Igel 2013; Castellanos *et al.* 2015; Zhao *et al.* 2016). The goal is to combine data from different events in one encoded ‘supergather’, thereby reducing the computational cost dramatically. The challenge with this approach had been that the encoded contributions from distinct sources are difficult to fully unravel, leading to ‘crosstalk’ (e.g. Romero *et al.* 2000), which contaminates the gradient of the misfit function. However, in recent years, successful methods of crosstalk-free source encoding have been developed by Huang & Schuster (2013, 2018), Zhang *et al.* (2018) and Krebs *et al.* (2013). In the approach of Zhang *et al.* (2018), encoded forward and adjoint wavefields are run until they reach steady state, at which point they are ‘deblended’ or ‘decoded’, based on an integration over a time interval that is the reciprocal of the encoded frequency spacing, to obtain their stationary parts. These parts are combined for all sources to obtain a crosstalk-free misfit gradient. Decoding is a critical requirement for the Zhang *et al.* (2018) approach to be successful, and this implementation differs from that of Huang & Schuster (2013, 2018) and Krebs *et al.* (2013), who invoke narrow-band filtering.

In earthquake seismology, Capdeville *et al.* (2005) proposed a method of source stacking that eliminates the linear dependence on the number of sources, but the approach requires that every source is recorded by all receivers. Unfortunately, this is basically never the case in practice. In exploration seismology this ‘fixed-receiver assumption’ was recently overcome (see e.g. Krebs *et al.* 2013; Dai *et al.* 2013; Huang & Schuster 2018).

The primary goal of this paper is to introduce source-encoding for earthquake seismology based on the (an)elastic wave equation for a wide variety of measurements with related misfit functions and adjoint sources. A challenge in this context is that not every earthquake is recorded by all receivers. A secondary goal is to present a time-domain version of the Zhang *et al.* (2018) source-encoding method that eliminates the need to store unnecessary intermediary products, such as the decoded components of the misfit gradient, thereby significantly reducing I/O and storage requirements. The resulting time-domain expression for the gradient of the misfit function is reminiscent of the time-domain reverse time migration (RTM) acoustic imaging condition of Dai *et al.* (2013). The time-domain approach has the additional benefit that it involves only minor modifications of the currently used adjoint tomography workflow. We attempt to be systematic and educational, with the aim of providing a complete overview of the method.

Throughout this paper, we use the Fourier convention of Dahlen & Tromp (1998, eq. 4.2), that is,

$$\mathbf{s}(\mathbf{x}, \omega) = \int_{-\infty}^{\infty} \mathbf{s}(\mathbf{x}, t) \exp(-i\omega t) dt, \quad (1)$$

with inverse

$$\mathbf{s}(\mathbf{x}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{s}(\mathbf{x}, \omega) \exp(i\omega t) d\omega. \quad (2)$$

Here \mathbf{s} denotes a vector displacement field, \mathbf{x} a position vector, t time and ω angular frequency.

2 FORWARD WAVEFIELD

In earthquake seismology, the seismic source may be represented as an equivalent body force of the form (Dahlen & Tromp 1998)

$$f_j(\mathbf{x}, t) = -M_{jk}(\mathbf{x}_s) \nabla_k \delta(\mathbf{x} - \mathbf{x}_s) S(t), \quad (3)$$

where M_{jk} denotes an element of centroid moment tensor, \mathbf{x}_s the centroid location, and $S(t)$ the source time function. In terms of the Green’s function, $G_{ij}(\mathbf{x}, \mathbf{x}', t - t')$, the displacement field due to a body force may be expressed in the form (Aki & Richards 1980; Dahlen & Tromp 1998)

$$s_i(\mathbf{x}, t) = \int_{-\infty}^t \int G_{ij}(\mathbf{x}, \mathbf{x}', t - t') f_j(\mathbf{x}', t') d^3 \mathbf{x}' dt'. \quad (4)$$

Written out explicitly, the displacement due to a point force (3) becomes

$$s_i(\mathbf{x}, t) = \int_{-\infty}^t \nabla_k G_{ij}(\mathbf{x}, \mathbf{x}_s; t - t') M_{kj}(\mathbf{x}_s) S(t') dt'. \quad (5)$$

Our goal will be to design a method for the calculation of Fréchet derivatives that combines a set of earthquakes into a single distributed encoded ‘super’ source, thereby significantly reducing the numerical cost of an inversion.

3 SOURCE ENCODING

Suppose we wish to perform adjoint tomography with a set of S sources in an angular frequency band $[\omega_{\min}, \omega_{\max}]$. The goal is to perform such an inversion based on just a single combination of encoded ‘super’ forward and adjoint simulations by effectively tagging each individual source. This is accomplished by randomly assigning each source s a unique frequency, ω_s , $s = 1, \dots, S$, defined by

$$\omega_s = \omega_{\min} + (s - 1) \Delta\omega, \quad (6)$$

where

$$\Delta\omega = (\omega_{\max} - \omega_{\min}) / (S - 1), \quad (7)$$

thereby covering the frequency band of interest, $[\omega_{\min}, \omega_{\max}]$.

Our approach will be to run encoded forward and adjoint simulations driven by a superposition of monochromatic sources with frequencies (6) until the wavefields reach steady state after some time T_{ss} . Following Zhang *et al.* (2018), the technique will take advantage of the fact that the source-encoded Fourier basis is orthogonal in the following sense:

$$\frac{1}{\Delta\tau} \int_{T_{ss}}^{T_{ss} + \Delta\tau} \exp(-i\omega_s t) \exp(i\omega_{s'} t) dt = \delta_{ss'}, \quad (8)$$

where

$$\Delta\tau = \frac{2\pi}{\Delta\omega} = \frac{2\pi(S - 1)}{\omega_{\max} - \omega_{\min}}. \quad (9)$$

We shall see that the time interval $\Delta\tau$ determines a period of integration required for ‘deblending’ or ‘decoding’ the encoded forward and adjoint wavefields, as well as the length of the convolution of the encoded forward and adjoint wavefields for the calculation of Fréchet derivatives.

While Huang & Schuster (2018) and Dai *et al.* (2013) choose $T_{ss} = \Delta\tau$, we believe that this choice leads to poor performance when many frequencies are encoded, or to strong spectral leakage in the gradient when few frequencies are encoded. On one hand, as expressed in eq. (9), $\Delta\tau$ solely depends on the available bandwidth and the number of encoded frequencies. On the other hand, T_{ss} is determined by simulation parameters, such as wave speeds and the spatial dimensions of the domain. Intuitively, steady state is obtained once transient effects disappear. Practically, one can compute the Fourier coefficients of the encoded wavefield evaluated at the encoded frequencies (6) over a time interval $[t_0; t_0 + \Delta\tau]$. There exists a time T_{ss} such that these coefficients remain unchanged for any $t_0 > T_{ss}$. This evaluation may be performed prior to inversion.

In practice, the set of frequencies determined by (6) is randomly distributed over the sources at the start of every iteration, and we demonstrate in the examples in Section 14 that one may choose to assign more than one frequency per source. If the number of frequencies per source is N_f , then there are a total of SN_f frequencies. In that case the bandwidth of interest is divided up accordingly:

$$\Delta\omega = (\omega_{\max} - \omega_{\min}) / (SN_f - 1), \quad (10)$$

and the time interval for integration becomes

$$\Delta\tau = \frac{2\pi(SN_f - 1)}{\omega_{\max} - \omega_{\min}}. \quad (11)$$

4 ENCODED FORWARD WAVEFIELD

To encode the forward wavefield, for a specific source s , we express eq. (5) in the frequency domain as

$$s_i^s(\mathbf{x}, \omega) = \nabla_k G_{ij}(\mathbf{x}, \mathbf{x}_s; \omega) M_{kj}(\mathbf{x}_s) S_s(\omega), \quad (12)$$

where S_s denotes the source–time function associated with source s . At source frequency ω_s , we define the stationary field

$$\begin{aligned} s_i^s(\mathbf{x}) &\equiv s_i^s(\mathbf{x}, \omega_s) \\ &= \nabla_k G_{ij}(\mathbf{x}, \mathbf{x}_s; \omega_s) M_{kj}(\mathbf{x}_s) S_s(\omega_s). \end{aligned} \quad (13)$$

Next, consider the following construct:

$$\begin{aligned}
S_i(\mathbf{x}, t) &\equiv \Re \sum_{s=1}^S s_i^s(\mathbf{x}) \exp(i\omega_s t) \\
&= \Re \sum_{s=1}^S \nabla_k G_{ij}(\mathbf{x}, \mathbf{x}_s; \omega_s) M_{kj}(\mathbf{x}_s) S_s(\omega_s) \exp(i\omega_s t) \\
&= \Re \sum_{s=1}^S M_{kj}(\mathbf{x}_s) S_s(\omega_s) \int_{-\infty}^{\infty} \nabla_k G_{ij}(\mathbf{x}, \mathbf{x}_s; t') \exp(-i\omega_s t') dt' \exp(i\omega_s t) \\
&= \Re \sum_{s=1}^S \int_{-\infty}^{\infty} \nabla_k G_{ij}(\mathbf{x}, \mathbf{x}_s; t - t') M_{kj}(\mathbf{x}_s) S_s(\omega_s) \exp(i\omega_s t') dt'.
\end{aligned} \tag{14}$$

Based on the results discussed in Section 2, the last equality motivates the definition of the encoded time-domain super source

$$F_j(\mathbf{x}, t) = \Re \sum_{s=1}^S f_j^s(\mathbf{x}, \omega_s) \exp(i\omega_s t), \tag{15}$$

where $f_j^s(\mathbf{x}, \omega_s)$ denotes the Fourier transform of body force (3) associated with source s . Effectively, all sources broadcast simultaneously, each at its own unique frequency ω_s . In Appendix A, we present encoded super sources for an earthquake modelled by a Gaussian moment-rate tensor and a shot modelled by a Ricker wavelet.

The encoded super source (15) gives rise to the encoded forward wavefield

$$S_i(\mathbf{x}, t) = \int_{-\infty}^t \int G_{ij}(\mathbf{x}, \mathbf{x}'; t - t') F_j(\mathbf{x}', t') d^3\mathbf{x}' dt'. \tag{16}$$

Thus the recipe for calculating the encoded super forward wavefield is simple: combine all earthquakes into the source-encoded super source (15) and run one forward simulation to obtain the encoded super forward wavefield (16).

5 DECODING THE ENCODED FORWARD WAVEFIELD

We run the encoded forward simulation (16) until the wavefield reaches steady state at a time T_{ss} , that is, until it can be expressed as a trigonometric polynomial

$$\begin{aligned}
S_i(\mathbf{x}, t) &= \Re \sum_{s=1}^S s_i^s(\mathbf{x}) \exp(i\omega_s t) \\
&= \sum_{s=1}^S [A_i^s(\mathbf{x}) \cos(\omega_s t) + B_i^s(\mathbf{x}) \sin(\omega_s t)].
\end{aligned} \tag{17}$$

Here we have defined

$$s_i^s(\mathbf{x}) = A_i^s(\mathbf{x}) - i B_i^s(\mathbf{x}), \tag{18}$$

where

$$A_i^s(\mathbf{x}) = \Re s_i^s(\mathbf{x}), \tag{19}$$

and

$$B_i^s(\mathbf{x}) = -\Im s_i^s(\mathbf{x}). \tag{20}$$

Using the orthogonality relationship (8), we may ‘deblend’ or ‘decode’ the stationary parts of the encoded forward wavefield $S_i(\mathbf{x}, t)$ based on the following integrations:

$$A_i^s(\mathbf{x}) = \frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} S_i(\mathbf{x}, t) \cos(\omega_s t) dt, \tag{21}$$

$$B_i^s(\mathbf{x}) = \frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} S_i(\mathbf{x}, t) \sin(\omega_s t) dt. \tag{22}$$

The time $\Delta\tau$ required for decoding is defined by eq. (9). Alternatively, the stationary parts may be determined based on trigonometric interpolation of the steady-state encoded forward wavefield using a nonuniform fast Fourier transform (Wright *et al.* 2015).

The Fourier coefficients of interest, $A_i^s(\mathbf{x})$ and $B_i^s(\mathbf{x})$, may also be obtained based on a discrete Fourier transform of the steady-state encoded forward wavefield (17). This is possible when the sampling frequency Δf of the seismograms is chosen such that (1) $\Delta\tau$ is a

multiple of $1/\Delta f$, (2) the Shannon theorem is respected, that is $\Delta f > \omega_{\max}/\pi$, and (3) $\omega_1 = \omega_{\min}$ must be chosen as a multiple of $\Delta\omega$. Under these conditions the vector generated by the fast Fourier transform has a subset that contains exactly the Fourier coefficients of the encoded frequencies ω_s , $s = 1, \dots, S$. The delay theorem must be applied to recover the correct phase at T_{ss} .

6 FULL WAVEFORM INVERSION

To develop intuition for inversions based on the encoded forward wavefield (16), we first consider the simple case of classical FWI.

6.1 Classical FWI

For a given shot or source s , the classical FWI misfit function is defined by (Tarantola 1984a, 1986)

$$\chi_s = \frac{1}{2} \sum_{r=1}^{R_s} \int [s_i^s(\mathbf{x}_r, t) - d_i^s(\mathbf{x}_r, t)]^2 dt, \quad (23)$$

where $d_i^s(\mathbf{x}_r, t)$ denote the data recorded at receiver r for source s , and where $r = 1, \dots, R_s$ labels the receivers that recorded event s . It is important to recognize that in earthquake seismology an event is typically recorded by a subset of the available receivers. The variation of the misfit function (23) is

$$\delta\chi_s = \sum_{r=1}^{R_s} \int [s_i^s(\mathbf{x}_r, t) - d_i^s(\mathbf{x}_r, t)] \delta s_i^s(\mathbf{x}_r, t) dt, \quad (24)$$

and the associated adjoint source is (Tromp *et al.* 2005)

$$f_j^{\dagger s}(\mathbf{x}, t) = \sum_{r=1}^{R_s} [s_i^s(\mathbf{x}_r, -t) - d_i^s(\mathbf{x}_r, -t)] \delta(\mathbf{x} - \mathbf{x}_r). \quad (25)$$

This adjoint source gives rise to the FWI adjoint wavefield

$$s_i^{\dagger}(\mathbf{x}, t) = \int_{-\infty}^t \int G_{ij}(\mathbf{x}, \mathbf{x}', t - t') f_j^{\dagger}(\mathbf{x}', t') d^3\mathbf{x}' dt'. \quad (26)$$

6.2 Source-encoded FWI

If the desired measurement is based on the Fourier coefficients of the full time-series of the seismograms, we Fourier transform the observed data $d_i^s(\mathbf{x}_r, t)$ for all sources $s = 1, \dots, S$, recorded by a set of receivers \mathbf{x}_r , $r = 1, \dots, R_s$, to obtain the Fourier coefficients $d_i^s(\mathbf{x}_r, \omega)$, which we evaluate at the encoded frequencies ω_s :

$$d_i^s(\mathbf{x}_r) \equiv d_i(\mathbf{x}_r, \omega_s). \quad (27)$$

To ensure that the discrete Fourier transform of the observed data contains exactly the encoded frequencies ω_s , $s = 1, \dots, S$, we may use resampling, such that the new sampling frequency respects the three conditions described at the end of Section 5, and zero padding, such that the integral

$$d_i^s(\mathbf{x}_r) = \frac{1}{k\Delta\tau} \int_0^{k\Delta\tau} d_i(\mathbf{x}_r, t) \exp(i\omega_s t) dt \quad (28)$$

can be computed. Here k is the smallest integer such that $k\Delta\tau > T$, where T denotes the length of the unpadded seismogram. It is worth noting that T and $\Delta\tau$ are not correlated. To obtain a good measurement, it is desirable that T captures the full time-series. It is not possible to perform any time windowing in seismograms prior to Fourier transformation, because these operations cannot be mimicked in an encoded simulation, that is windowing requires simulations of individual events.

Next, we consider the encoded waveform misfit function

$$\chi = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} [s_i^{s*}(\mathbf{x}_r) - d_i^{s*}(\mathbf{x}_r)][s_i^s(\mathbf{x}_r) - d_i^s(\mathbf{x}_r)]. \quad (29)$$

Effectively this misfit function is a frequency-domain version of (23) summed over all sources. In Section 14, we will discuss the options of randomizing the frequency assigned to each source, as well as the option of assigning multiple frequencies to the same source. The variation

of this encoded misfit function may be expressed as

$$\begin{aligned}
\delta\chi &= \Re \sum_{s=1}^S \sum_{r=1}^{R_s} [s_i^{s*}(\mathbf{x}_r) - d_i^{s*}(\mathbf{x}_r)] \delta s_i^s(\mathbf{x}_r) \\
&= \Re \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{s'=1}^S [s_i^{s*}(\mathbf{x}_r) - d_i^{s*}(\mathbf{x}_r)] \delta_{ss'} \delta s_i^{s'}(\mathbf{x}_r) \\
&= \Re \frac{1}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} \sum_{s=1}^S \sum_{r=1}^{R_s} [s_i^{s*}(\mathbf{x}_r) - d_i^{s*}(\mathbf{x}_r)] \exp(-i\omega_s t) \sum_{s'=1}^S \delta s_i^{s'}(\mathbf{x}_r) \exp(i\omega_{s'} t) dt \\
&= \Re \frac{1}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} \int \sum_{s=1}^S \sum_{r=1}^{R_s} [s_i^{s*}(\mathbf{x}_r) - d_i^{s*}(\mathbf{x}_r)] \delta(\mathbf{x} - \mathbf{x}_r) \exp(-i\omega_s t) \sum_{s'=1}^S \delta s_i^{s'}(\mathbf{x}) \exp(i\omega_{s'} t) d^3\mathbf{x} dt.
\end{aligned} \tag{30}$$

where we used the orthogonality (8) in the third equality.

Using a technique for identifying the adjoint source that corresponds to the misfit variation (30) explained in Bozdağ *et al.* (2011, Section 2.1), we find that it is given by

$$F_j^\dagger(\mathbf{x}, t) = \Re \sum_{s=1}^S f_j^{\dagger s}(\mathbf{x}, \omega_s) \exp(i\omega_s t), \tag{31}$$

where $f_j^{\dagger s}(\mathbf{x}, \omega_s)$ denotes the Fourier transform of the classical waveform adjoint source (25) associated with source s . To avoid spectral leakage, we emphasize that it is critical that the Fourier coefficients $d_i^s(\mathbf{x}_r)$ and $s_i^s(\mathbf{x}_r)$ are evaluated *exactly* at the source encoded frequencies, not approximated via a narrow bandpass filter (as in, e.g. Huang & Schuster 2012, 2018).

It is important to note that the encoded super adjoint source (31) may actually be constructed directly from the stationary parts of the encoded forward wavefield, $s_i^s(\mathbf{x}_r)$, and the data Fourier coefficients (27), $d_i^s(\mathbf{x}_r)$. In other words, the waveform adjoint source may be calculated based on a ‘super measurement’, requiring just a single super adjoint calculation.

The ‘super’ adjoint wavefield $S_i^\dagger(\mathbf{x}, t)$ generated by the encoded adjoint source (31) may be expressed in the form

$$S_i^\dagger(\mathbf{x}, t) = \int_{-\infty}^t \int G_{ij}(\mathbf{x}, \mathbf{x}'; t - t') F_j^\dagger(\mathbf{x}', t') d^3\mathbf{x}' dt'. \tag{32}$$

Thus, like the encoded super forward wavefield (16), the recipe for calculating the encoded super adjoint wavefield is simple: combine all the usual adjoint sources into the source-encoded super adjoint source (31) and run one simulation to obtain the encoded super adjoint wavefield (32).

In anticipation of the Fréchet derivatives discussed in Section 10, we note that the combined sums over s and r in the misfit variation (30) give rise to the encoded adjoint wavefield (32), whereas the sum over s' gives rise to the encoded forward wavefield (16).

7 PHASE, AMPLITUDE, TRAVELTIME AND ‘DOUBLE-DIFFERENCE’ MISFIT FUNCTIONS

The choice of measurement and related misfit function is a critical component of a successful inversion strategy. In the interest of succinctness, we have relegated discussions of source-encoded misfit functions and related adjoint sources involving phase, amplitude, cross-correlation traveltimes and various ‘double-difference’ (Yuan *et al.* 2016) measurements to Appendices B and C. We show that phase, amplitude, and even hybrid (phase + amplitude) measurements are amenable to double-differencing, as opposed to conventional waveform measurements. Double-difference measurements hold great promise for source-encoded adjoint tomography, because they do not require any knowledge of the original source–time function or the instrument response.

8 SOURCE-ENCODED ADJOINT TOMOGRAPHY

The results obtained in Section 6.2 for FWI—and in Appendices B and C for phase, amplitude, and cross-correlation traveltimes—may be generalized as follows. Suppose we have a set of adjoint sources $f_j^{\dagger s}(\mathbf{x}, t)$, $s = 1, \dots, S$, one for each earthquake in the database. These adjoint sources are constructed in the usual fashion based on measurements of the differences between observed and simulated seismograms. To obtain the encoded adjoint wavefield, the classical waveform adjoint sources are combined into an encoded adjoint super source

$$F_j^\dagger(\mathbf{x}, t) = \Re \sum_{s=1}^S f_j^{\dagger s}(\mathbf{x}, \omega_s) \exp(i\omega_s t). \tag{33}$$

Based on this source, the super adjoint wavefield is given by

$$S_i^\dagger(\mathbf{x}, t) = \int_{-\infty}^t \int G_{ij}(\mathbf{x}, \mathbf{x}'; t - t') F_j^\dagger(\mathbf{x}', t') d^3\mathbf{x}' dt'. \tag{34}$$

9 DECODING THE ENCODED ADJOINT WAVEFIELD

We run the super adjoint wavefield simulation (34) until it reaches steady state at time T_{ss} . At that stage we express it in the form

$$\begin{aligned} S_i^\dagger(\mathbf{x}, t) &= \sum_{s=1}^S [A_i^{\dagger s}(\mathbf{x}) \cos(\omega_s t) - B_i^{\dagger s}(\mathbf{x}) \sin(\omega_s t)] \\ &= \Re \sum_{s=1}^S s_i^{\dagger s}(\mathbf{x}) \exp(-i\omega_s t), \end{aligned} \quad (35)$$

where

$$s_i^{\dagger s}(\mathbf{x}) = A_i^{\dagger s}(\mathbf{x}) - i B_i^{\dagger s}(\mathbf{x}), \quad (36)$$

and where the stationary parts

$$A_i^{\dagger s}(\mathbf{x}) = \Re s_i^{\dagger s}(\mathbf{x}), \quad (37)$$

and

$$B_i^{\dagger s}(\mathbf{x}) = -\Im s_i^{\dagger s}(\mathbf{x}), \quad (38)$$

may be decoded based on the expressions

$$A_i^{\dagger s}(\mathbf{x}) = \frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} S_i^\dagger(\mathbf{x}, t) \cos(\omega_s t) dt, \quad (39)$$

$$B_i^{\dagger s}(\mathbf{x}) = \frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} S_i^\dagger(\mathbf{x}, t) \sin(\omega_s t) dt. \quad (40)$$

10 FRÉCHET DERIVATIVES

For an isotropic medium, using the adjoint state method (for a review, see Plessix 2006), the variation in the encoded misfit function may be expressed in the form (Tromp *et al.* 2005)

$$\delta\chi = \int (\delta \ln \rho K_\rho + \delta \ln \kappa K_\kappa + \delta \ln \mu K_\mu) d^3\mathbf{x}, \quad (41)$$

where K_ρ , K_κ and K_μ denote Fréchet derivatives with respect to density ρ , bulk modulus κ , and shear modulus μ , respectively. Based on the encoded forward and adjoint wavefields $S_i(\mathbf{x}, t)$ and $S_i^\dagger(\mathbf{x}, t)$, these Fréchet derivatives are determined as follows, using the density derivative as an example:

$$K_\rho(\mathbf{x}) = -\frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} \rho(\mathbf{x}) S_i^\dagger(\mathbf{x}, -t) \partial_t^2 S_i(\mathbf{x}, t) dt \quad (42)$$

$$= \sum_{s=1}^S \omega_s^2 \rho(\mathbf{x}) [A_i^{\dagger s}(\mathbf{x}) A_i^s(\mathbf{x}) + B_i^{\dagger s}(\mathbf{x}) B_i^s(\mathbf{x})] \quad (43)$$

$$= \Re \sum_{s=1}^S \omega_s^2 \rho(\mathbf{x}) s_i^{\dagger s*}(\mathbf{x}) s_i^s(\mathbf{x}). \quad (44)$$

Note that there is no cross talk between sources, thanks to the orthogonality (8) of the constituent monochromatic waves. It is important to recognize that one may either calculate the density Fréchet derivative by convolving the encoded forward and adjoint wavefields based on the equality (42), effectively a time-domain approach, or one may choose to calculate it by summing the decoded contributions for each source based on either (43) or (44), effectively a frequency-domain approach. In Section 12, we present two complementary recipes for the calculation of Fréchet derivatives based on these time- and frequency-domain approaches.

Similarly, the remaining Fréchet derivatives are

$$K_\kappa(\mathbf{x}) = -\frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} \kappa(\mathbf{x}) [\nabla_i S_i^\dagger(\mathbf{x}, -t)] [\nabla_j S_j(\mathbf{x}, t)] dt \quad (45)$$

$$= -\sum_{s=1}^S \kappa(\mathbf{x}) \{[\nabla_i A_i^{\dagger s}(\mathbf{x})] [\nabla_j A_j^s(\mathbf{x})] + [\nabla_i B_i^{\dagger s}(\mathbf{x})] [\nabla_j B_j^s(\mathbf{x})]\} \quad (46)$$

$$= -\Re \sum_{s=1}^S \kappa(\mathbf{x}) [\nabla_i s_i^{\dagger s*}(\mathbf{x})] [\nabla_j s_j^s(\mathbf{x})], \quad (47)$$

$$K_\mu(\mathbf{x}) = -\frac{2}{\Delta\tau} \int_{T_{ss}}^{T_{ss}+\Delta\tau} 2\mu(\mathbf{x}) D_{ij}^\dagger(\mathbf{x}, -t) D_{ij}(\mathbf{x}, t) dt \quad (48)$$

$$= -\sum_{s=1}^S 2\mu(\mathbf{x}) \left\{ \left[\frac{1}{2}(\nabla_i A_j^{\dagger s} + \nabla_j A_i^{\dagger s}) - \frac{1}{3} \nabla_k A_k^{\dagger s} \delta_{ij} \right] \left[\frac{1}{2}(\nabla_i A_j^s + \nabla_j A_i^s) - \frac{1}{3} \nabla_k A_k^s \delta_{ij} \right] \right. \\ \left. + \left[\frac{1}{2}(\nabla_i B_j^{\dagger s} + \nabla_j B_i^{\dagger s}) - \frac{1}{3} \nabla_k B_k^{\dagger s} \delta_{ij} \right] \left[\frac{1}{2}(\nabla_i B_j^s + \nabla_j B_i^s) - \frac{1}{3} \nabla_k B_k^s \delta_{ij} \right] \right\} \quad (49)$$

$$= -\Re \sum_{s=1}^S 2\mu(\mathbf{x}) D_{ij}^{\dagger s*}(\mathbf{x}) D_{ij}^s(\mathbf{x}), \quad (50)$$

where $D_{ij}(\mathbf{x}, t)$ and $D_{ij}^\dagger(\mathbf{x}, t)$ denote the strain deviators associated with the encoded forward and adjoint wavefields, respectively, and where

$$D_{ij}^s = \frac{1}{2}(\nabla_i s_j^s + \nabla_j s_i^s) - \frac{1}{3} \nabla_k s_k^s \delta_{ij}, \quad (51)$$

and

$$D_{ij}^{\dagger s} = \frac{1}{2}(\nabla_i s_j^{\dagger s} + \nabla_j s_i^{\dagger s}) - \frac{1}{3} \nabla_k s_k^{\dagger s} \delta_{ij}, \quad (52)$$

are the corresponding decoded contributions. Further extensions to anisotropic media are straightforward (e.g. Tromp *et al.* 2005).

11 ATTENUATION

Interestingly, when using decoding, the calculation of the shear Q kernel no longer requires an additional adjoint simulation. The adjoint source for attenuation may be obtained from the regular adjoint source via the transformation (Tromp *et al.* 2005)

$$f_{Qj}^\dagger(\mathbf{x}, \omega) = [(2/\pi) \ln(|\omega|/\omega_0) - i \operatorname{sgn}(\omega)] f_j^\dagger(\mathbf{x}, \omega). \quad (53)$$

Consequently, the Q_μ Fréchet derivative is simply a modification of the K_μ kernel (50), namely,

$$K_{Q_\mu}(\mathbf{x}) = -\Re \sum_{s=1}^S [(2/\pi) \ln(|\omega_s|/\omega_0) - i \operatorname{sgn}(\omega_s)] 2\mu(\mathbf{x}) D_{ij}^{\dagger s*}(\mathbf{x}) D_{ij}^s(\mathbf{x}). \quad (54)$$

If we want to avoid decoding, we do need to calculate a second encoded adjoint wavefield, $S_{Qj}^\dagger(\mathbf{x}, t)$, driven by

$$F_{Qj}^\dagger(\mathbf{x}, t) = \Re \sum_{s=1}^S [(2/\pi) \ln(|\omega_s|/\omega_0) - i \operatorname{sgn}(\omega_s)] f_j^\dagger(\mathbf{x}, \omega_s) \exp(i\omega_s t). \quad (55)$$

12 RECIPE

The calculation of Fréchet derivatives is initiated by combining all earthquakes or shots into a single source-encoded super source (15), followed by simulating the corresponding source-encoded super forward wavefield (16) until it reaches steady state. The next ingredient involves calculation of the source-encoded super adjoint wavefield (34) until it reaches steady state. This requires the construction of the source-encoded super adjoint source (33). In this context, in adjoint tomography measurements are made by comparing observed and simulated seismograms in carefully selected time windows, a process that requires the calculation of individual synthetic seismograms for all sources at a cost that scales linearly with the number of sources. In contrast, in classical FWI adjoint sources may be constructed based on the source-encoded Fourier coefficients of the observed and simulated wavefields, effectively involving a single ‘super measurement’. The latter may be obtained directly from the source-encoded super forward wavefield, therefore not requiring the calculation of individual synthetic seismograms. Thus, desirably and remarkably, one iteration in the inverse problem requires just two simulations, independent of the number of sources or receivers.

Once the super forward and adjoint wavefields have been calculated, source-encoded Fréchet derivatives may be obtained based on one of two mathematically equivalent approaches, one in the time domain and the other in the frequency domain. These two approaches lead to the following options for the calculation of Fréchet derivatives.

12.1 Option I: frequency-domain approach

In the first approach, steady-state, source-encoded forward and adjoint simulations are decoded to obtain a set of stationary fields proportional to the number of sources in the dataset (eqs 21, 22, 39 and 40). These stationary parts are subsequently combined to calculate Fréchet derivatives by summing their respective contributions (eqs 43, 46 and 49). In the context of classical FWI, this option is basically the approach first proposed by Zhang *et al.* (2018). An advantage of this option is that the encoded forward and adjoint wavefields may be calculated simultaneously and independently. Another advantage is that attenuation poses no challenges, because both wavefields are used only in

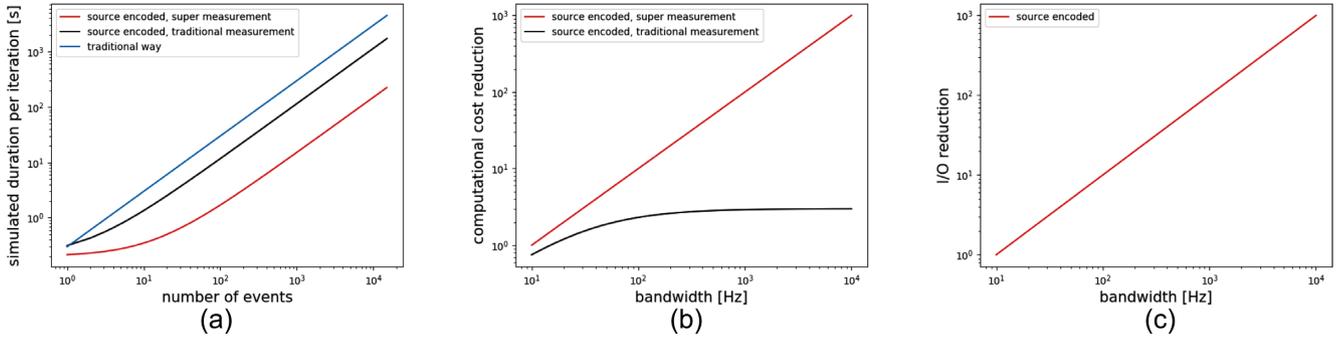


Figure 1. Potential speed-up depending on measurement type and allocated frequency bandwidth. We assume a simulation per event of duration $T = 0.1$ s and a bandwidth of 200 Hz, which reflects the model setup shown in Fig. 3. As a reasonable assumption, we set $T_{ss} \simeq T$. (a) Simulation costs for traditional inversions (56), source-encoded inversions (57), and source-encoded inversions with traditional measurements (59) (blue, red, and black curves, respectively). Because super measurements are possible, one can expect a 20X speed-up in this configuration. (b) Computational cost reduction relative to a traditional inversion for source-encoded inversions (58) and source-encoded inversions with traditional measurements (60) (red and black curves, respectively), that is, the difference between the blue curve and the other two curves shown in (a). We see the expected computational cost reduction from source encoding, which corresponds to the asymptotic behavior of the curves in (a). Because time domain measurements require an individual simulation for each event, no speed-up bigger than a factor of 3 can be expected. (c) I/O reduction based on eq. (63). Whether conventional or super measurements are considered, I/O is drastically reduced when considering a larger bandwidth.

forward time. A disadvantage of this option is that the stationary parts of the encoded wavefields must be computed simultaneously. This requires significant RAM, even assuming that Fourier transforms are performed on the fly and that no time storage of the wavefields is necessary. Each stationary part consists of a 3-D volume of displacement, and the number of such volumes scales linearly with the number of sources, which can run in the thousands. Ultimately, we are only interested in obtaining Fréchet derivatives, and neither the stationary parts of the wavefields nor their contributions to the Fréchet derivatives are of any practical value.

12.2 Option II: time-domain approach

In the second approach, steady-state source-encoded forward and adjoint wavefields are simply convolved over a time period proportional to the inverse of the frequency bandwidth of interest, thereby directly rendering the Fréchet derivatives (eqs 42, 45 and 48). This approach is reminiscent of the time-domain source-encoded RTM acoustic imaging condition derived by Dai *et al.* (2013), although they do not distinguish between the time it takes to reach steady state, T_{ss} , and the integration interval required for convolution, $\Delta\tau$. In this option there is no need to decode the encoded forward and adjoint wavefields, nor do intermediary contributions to the misfit gradient need to be calculated or stored. Attractively, this option requires only minor modifications of the currently used adjoint tomography workflow. As in the current workflow, the encoded forward wavefield must be reconstructed during the calculation of the encoded adjoint wavefield so that they may be convolved, a challenge elegantly solved by the wavefield reconstruction algorithm of Komatitsch *et al.* (2016). It is important to note that in the source-encoded approach only the super forward wavefield needs to be reconstructed during the calculation of the Fréchet derivatives, whereas in the classical approach this procedure must be executed for every single forward wavefield, that is, for every event in the database. This translates into a significantly smaller amount of I/O, independent of the number of events.

13 COMPARISON OF COMPUTATIONAL COSTS

Traditional time-domain FWI or adjoint tomography requires one forward simulation of a chosen duration T per event to capture the seismic arrivals of interest, and one adjoint calculation of the same duration to determine the contribution of this event to the gradient. Unless the forward calculation can be stored on disk, the adjoint calculation is twice as expensive as the forward simulation, because the forward wavefield needs to be reconstructed during the adjoint simulation for convolution. In this context, many industrial applications use checkpointing (see e.g. Anderson *et al.* 2012). Thus, for a given event, cumulative seismograms are calculated for a total duration of $3T$, and a total of

$$T_{\text{traditional}} = 3TS \quad (56)$$

simulation time is required for one iteration involving S events.

In source-encoded FWI, one super forward calculation of duration $T_{ss} + \Delta\tau$ is required, followed by one super adjoint calculation of the same duration. The transient part of the forward wavefield, T_{ss} , does not need to be reconstructed to compute the sensitivity kernels. Thus,

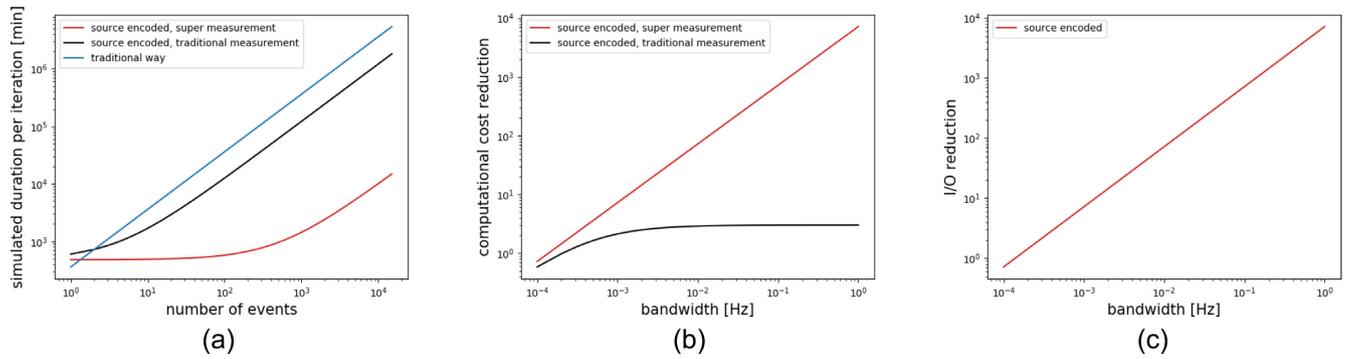


Figure 2. Potential speed-up for global FWI. In this case we assume $T_{ss} \simeq 2T$. (a) Simulation costs for traditional inversions (56), source-encoded inversions (57), and source-encoded inversions with traditional measurements (59) (blue, red, and black curves, respectively) with a bandwidth of 0.05 Hz. If super measurements can be used, we would see a 360X reduction in cost. For a bandwidth of 1 Hz, we would see a 7200X reduction in cost. (b) Computational cost reduction relative to a traditional global FWI inversion for source-encoded inversions (58) and source-encoded inversions with traditional measurements (60) (red and black curves, respectively), that is, the difference between the blue curve and the other two curves shown in (a). We see that the best we can do if we make traditional measurements, which requires individual simulations for each event, is a 3X reduction in simulation time. (c) I/O reduction based on eq. (63). We see that the I/O reduction for source-encoded global FWI is significant at 360X, regardless of whether individual measurements are made.

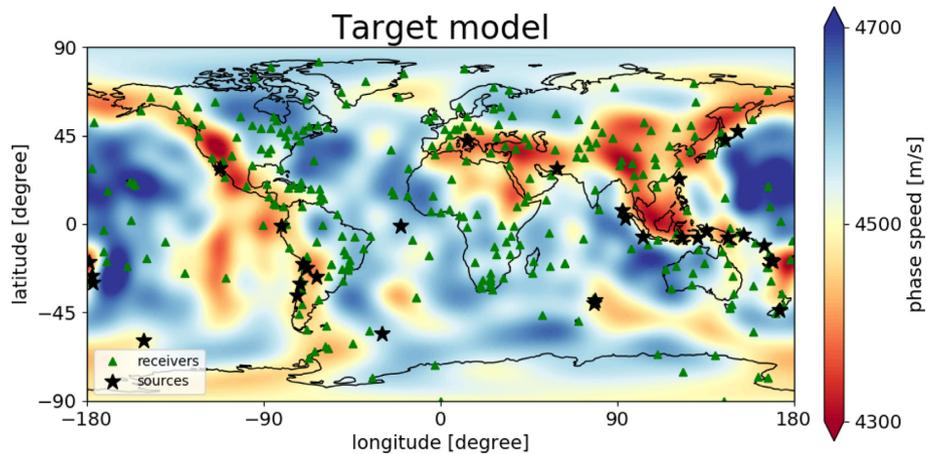


Figure 3. Target model used to illustrate inversion options and strategies. The model is a 50 s Rayleigh wave phase speed map determined by Trampert & Woodhouse (2003). There are 32 sources labelled by black stars and 293 seismographic stations labelled by green triangles.

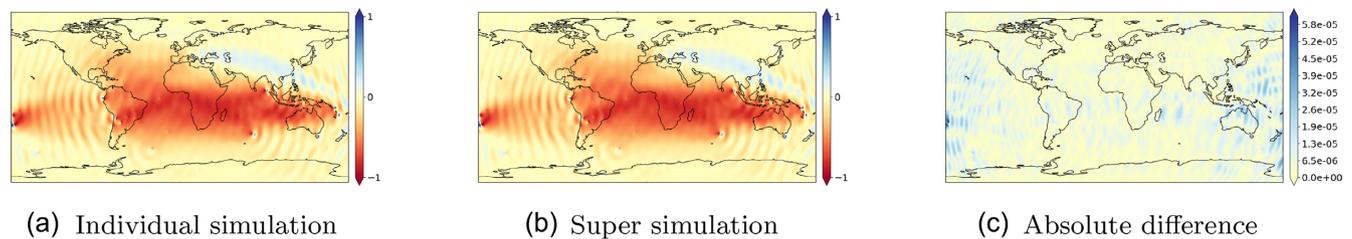


Figure 4. Illustration of the absence of cross talk between encoded sources in the frequency domain. (a) Conventional approach in which sensitivity kernels (“event kernels”) are calculated individually in the frequency domain for each event and summed to obtain the gradient of the misfit function. This process involves 64 simulations, two for each earthquake in the database. (b) Source-encoded approach in which the gradient is calculated in the frequency domain based on eq. (50) using one super forward and one super adjoint simulation, that is, two simulations, independent of the number of sources and receivers. Panels (a) and (b) are rescaled in the range $[-1, +1]$ using a common factor. (c) Absolute value of the difference between (a) and (b). The largest discrepancy between the two gradients over the entire domain is 0.0058 per cent.

using eq. (9), a total simulation time of

$$\begin{aligned}
 T_{\text{encoded}} &= \underbrace{2(T_{ss} + \Delta\tau)}_{\text{super forward and adjoint}} + \underbrace{\Delta\tau}_{\text{steady super forward reconstruction}} \\
 &= 2T_{ss} + 3\Delta\tau \\
 &= 2T_{ss} + \frac{3(S-1)}{f_{\text{max}} - f_{\text{min}}}
 \end{aligned} \tag{57}$$

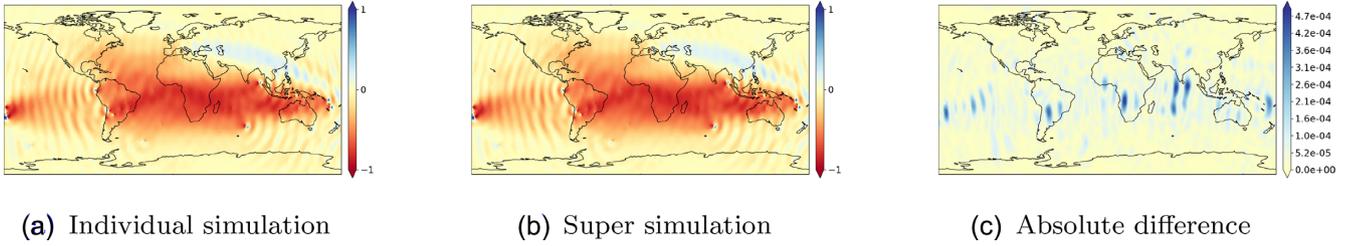


Figure 5. Illustration of the absence of cross talk between encoded sources in the time domain. (a) Traditional approach; same as Fig. 4(a), but with kernel computation in the time domain. (b) Source-encoded approach in which the gradient is calculated in the time domain based on eq. (48). The two gradients are rescaled in the range $[-1, +1]$ using a common factor. The small differences shown in (c) prove that frequency-domain decoding is not necessary to compute the super sensitivity kernel, thanks to the mathematical equivalence of eqs (48) and (50). The largest discrepancy between the two gradients over the entire domain is 0.048 per cent.

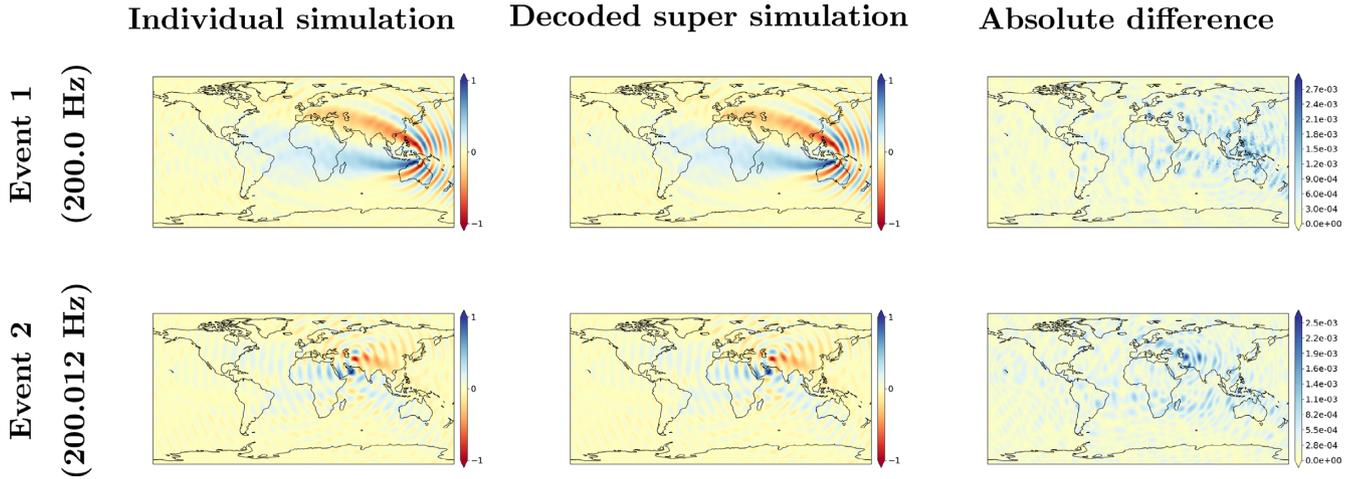


Figure 6. Comparison of the contribution of a single frequency to the overall gradient calculated individually (left-hand column) and calculated simultaneously and decoded (middle column). During the super simulation, each of the 32 events is encoded with 512 frequencies, for a total of 16 384 frequencies over the $[200; 400]$ Hz bandwidth, using a frequency spacing of 0.01207 Hz. We show the two first encoded frequencies (rows 1 and 2), each belonging to a different event, to illustrate the absence of crosstalk between two very close frequencies. For the super simulation, each contribution was recovered using decoding as defined by eqs (39) and (40). All gradients are rescaled in the range $[-1, +1]$ using a common factor for each row. The very small differences between the individual and super simulations (right-hand column) provide further proof of the absence of cross talk between sources during the simultaneous run.

is required per source-encoded iteration. The relative reduction in computational cost is

$$\frac{T_{\text{traditional}}}{T_{\text{encoded}}} = \frac{3TS}{2T_{\text{ss}} + \frac{3(S-1)}{f_{\text{max}} - f_{\text{min}}}} \quad (58)$$

$$\simeq T(f_{\text{max}} - f_{\text{min}}),$$

where the last equation holds asymptotically for large values of S , that is for a large number of events.

If we need to make measurements in the traditional way before calculating the misfit gradient based on source encoding, then we need to add T simulation time per event, and thus the cumulative simulation time becomes

$$T_{\text{encoded+measurements}} = \underbrace{TS}_{\text{measurements}} + \underbrace{2(T_{\text{ss}} + \Delta\tau)}_{\text{super forward and adjoint}} + \underbrace{\Delta\tau}_{\text{steady super forward reconstruction}} \quad (59)$$

$$= TS + 2T_{\text{ss}} + \frac{3(S-1)}{f_{\text{max}} - f_{\text{min}}}.$$

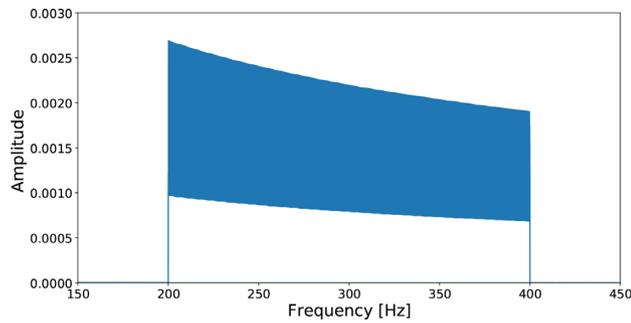
In this case the relative reduction in computational cost is

$$\frac{T_{\text{traditional}}}{T_{\text{encoded+measurements}}} = \frac{3TS}{TS + 2T_{\text{ss}} + \frac{3(S-1)}{f_{\text{max}} - f_{\text{min}}}} \quad (60)$$

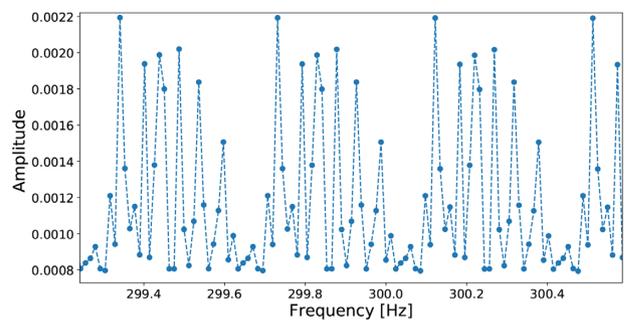
$$\simeq \frac{1}{\frac{1}{3} + \frac{1}{T(f_{\text{max}} - f_{\text{min}})}},$$

where the last equation holds asymptotically for large values of S . We conclude that in this case the best we can expect is a 3X speed-up.

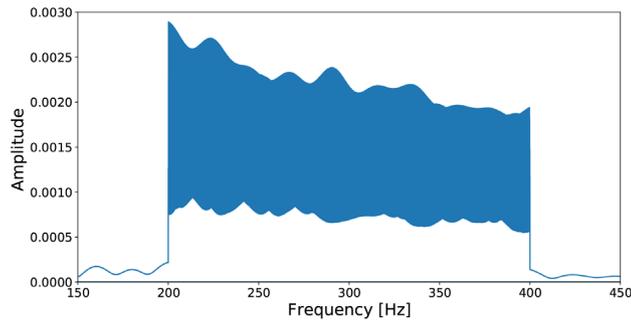
With regards to read/write operations (I/O), to calculate traditional sensitivity kernels one can (1) store the entire forward wavefield, (2) attempt to reconstruct the forward wavefield with absorbing boundary storage or (3) reconstruct the forward wavefield on the fly based on



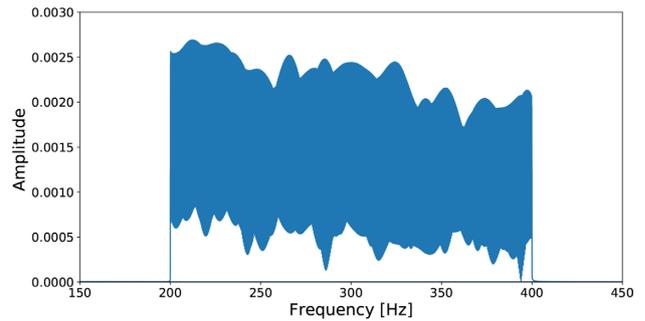
(a) Correctly decoded seismogram.



(b) Subinterval of (a).



(c) Incorrect time interval when decoding.



(d) Transient state embedded in decoding.

Figure 7. Study of the impact of approximations on decoding. We use a randomly selected seismogram from the experiment shown in Fig. 6, involving a super simulation encoding 16.384 frequencies in the [200;400] Hz range. The simulation lasts 1 640 400 time steps, including 2000 time steps of a transient state (defining T_{ss}) and 1 638 400 time steps of steady state (defining $\Delta\tau$). (a) Correctly decoded seismogram. We notice the progressive decrease of the Fourier amplitude with increasing frequency, which matches the theoretical expectation $A(\omega) \propto 1/\sqrt{\omega}$. (b) Zoom in on a narrow frequency subinterval of (a) to observe individual contributions of each event. Every event uses 512 frequencies, spaced such that the same event is encoded every 32 consecutive frequencies. As expected, we obtain a periodic signal (repeating itself every 32 samples) and clearly distinguish the contribution of each event, further demonstrating orthogonality. (c) An (incorrect) time interval of 1 638 390 time steps is used to decode, which leads to failure, because the wrong frequencies are evaluated and the orthogonality condition (8) is not satisfied. (d) Decoding performed over a time interval containing 1000 time steps of the transient state. Because the evaluated signal is no longer a pure trigonometric polynomial, orthogonality is lost, also resulting in incorrect decoding.

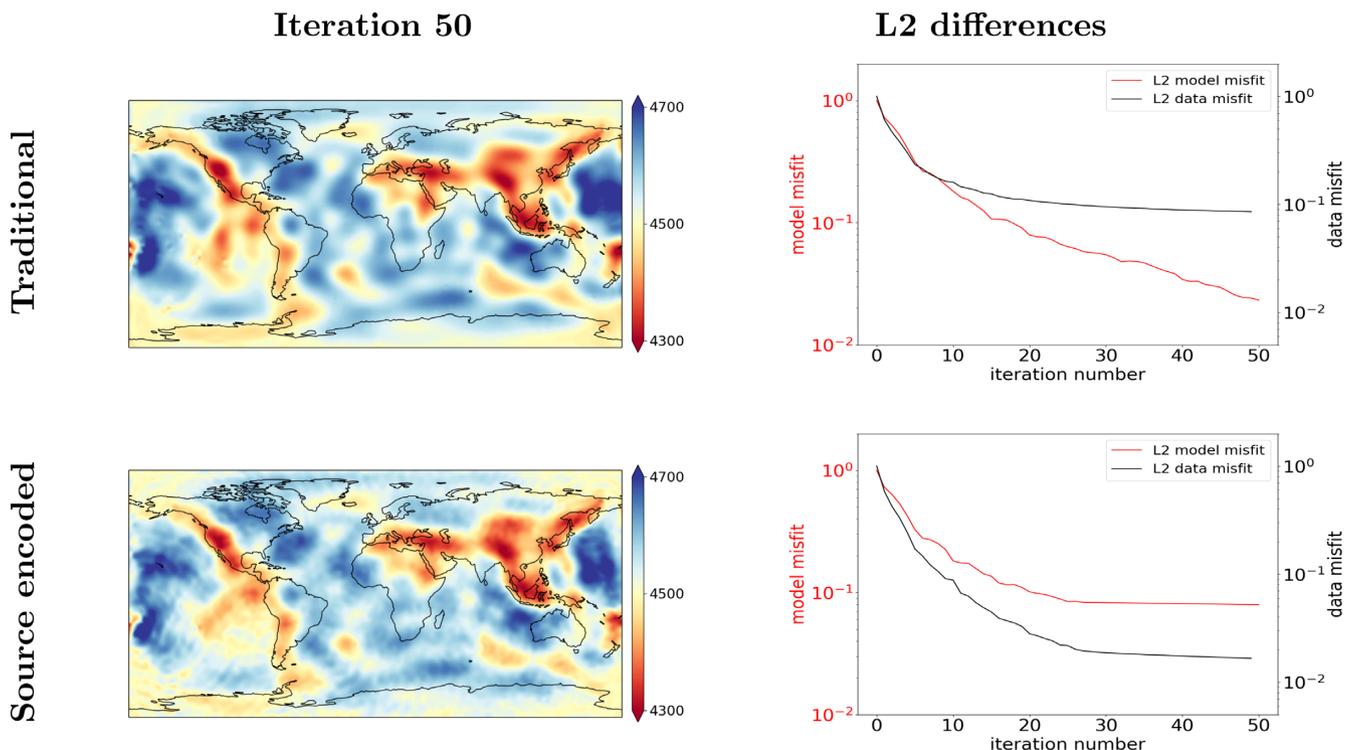


Figure 8. Comparison between the usual full spectrum inversion and the source encoded inversion. The encoded version uses 32 frequencies, one per event, in the [200;400] Hz bandwidth. Global convergence is achieved in both cases, even if some extra oscillations can be observed in the source encoded version. This is due to the drastic reduction of the frequency content of the inverted data, which can be mitigated as illustrated in Figs 9 and 10.

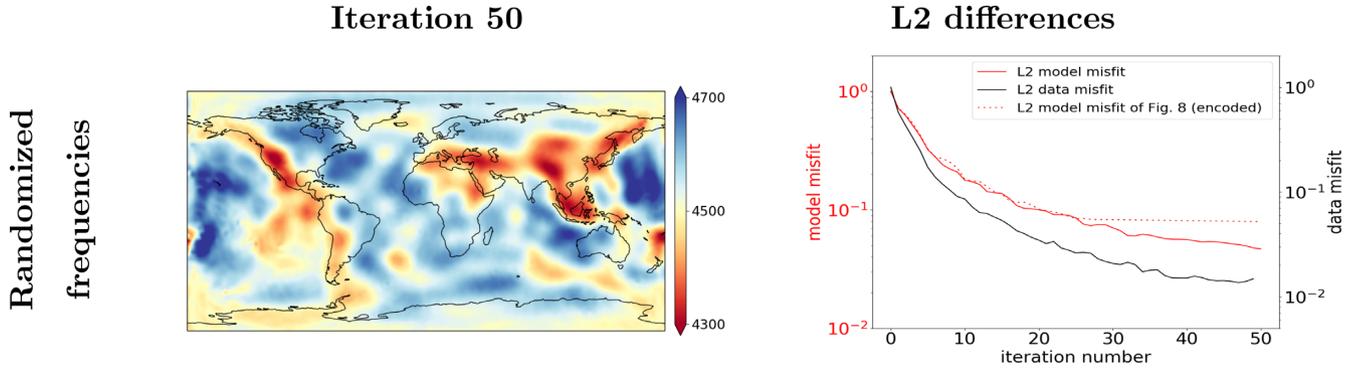


Figure 9. Same set-up as Fig. 8, but using randomization of the frequency content. Instead of attributing the same frequency to an event for the entire inversion, the frequency is changed at each iteration. This allows assimilation of more of the data content into the result, but also helps to better balance the overall sensitivity. This translates into an increased convergence rate. The computational cost remains unchanged.

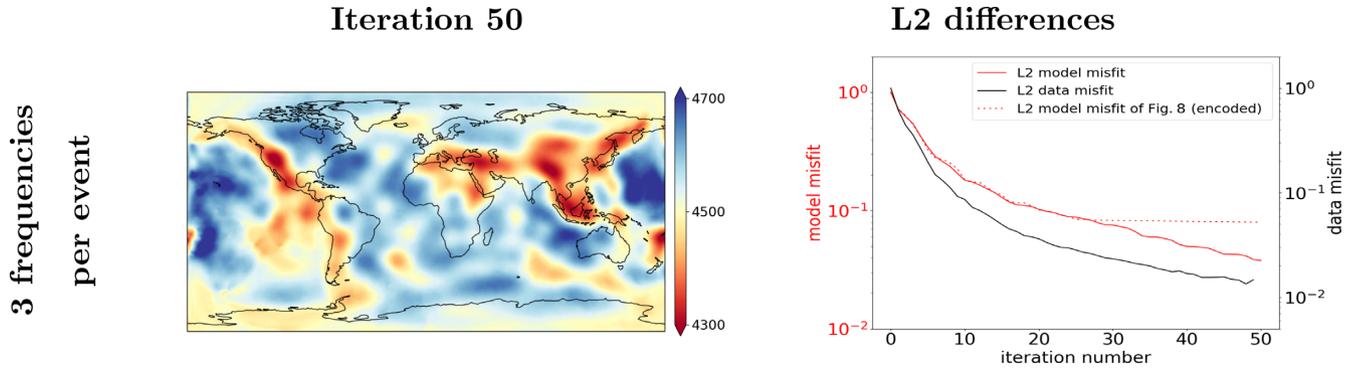


Figure 10. Same set-up as Fig. 9, but using a [150;750] Hz bandwidth, which enables encoding of 3 frequencies per event at the same computational cost. The increased amount of inverted data per iteration and the larger bandwidth of inverted data translates into faster convergence in model space.

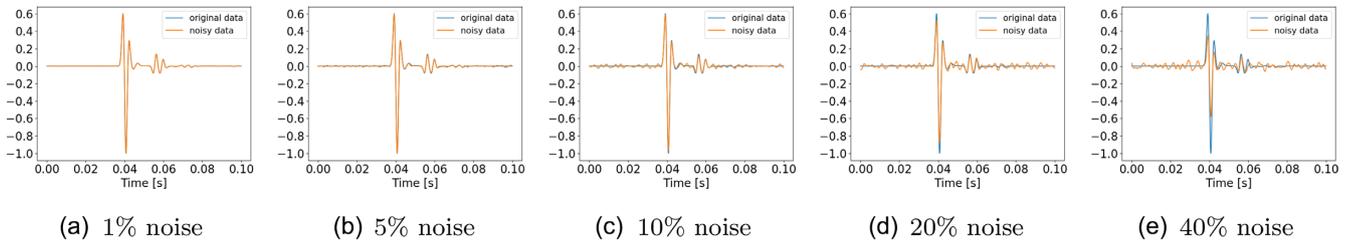


Figure 11. Example of noisy data used for inversion in Fig. 12. A Gaussian perturbation in phase is added to each frequency in the [50;750] Hz range, with zero mean and a standard deviation of $2\pi i$, where i is the percentage denoted for each row.

the parsimonious storage algorithm of Komatitsch *et al.* (2016). In any case, I/O is proportional to the cumulative forward simulation time, $T_{\text{traditional}}$. For large meshes, the remaining I/O costs (seismogram writing, adjoint source reading, sensitivity kernel writing, etc.) become negligible. Thus we have

$$I/O_{\text{traditional}} = \alpha T_{\text{traditional}}, \quad (61)$$

for some suitably chosen proportionality factor α .

For a source encoded simulation, I/O is only required for the sensitivity kernel computation, and does not involve any storage or reconstruction of the transient part of the super forward wavefield. Thus I/O for an encoded simulation is determined by

$$I/O_{\text{encoded}} = \alpha \frac{3(S-1)}{f_{\text{max}} - f_{\text{min}}}, \quad (62)$$

and the relative reduction in I/O is

$$\begin{aligned} \frac{I/O_{\text{traditional}}}{I/O_{\text{encoded}}} &= \frac{TS(f_{\text{max}} - f_{\text{min}})}{S-1} \\ &\simeq T(f_{\text{max}} - f_{\text{min}}). \end{aligned} \quad (63)$$

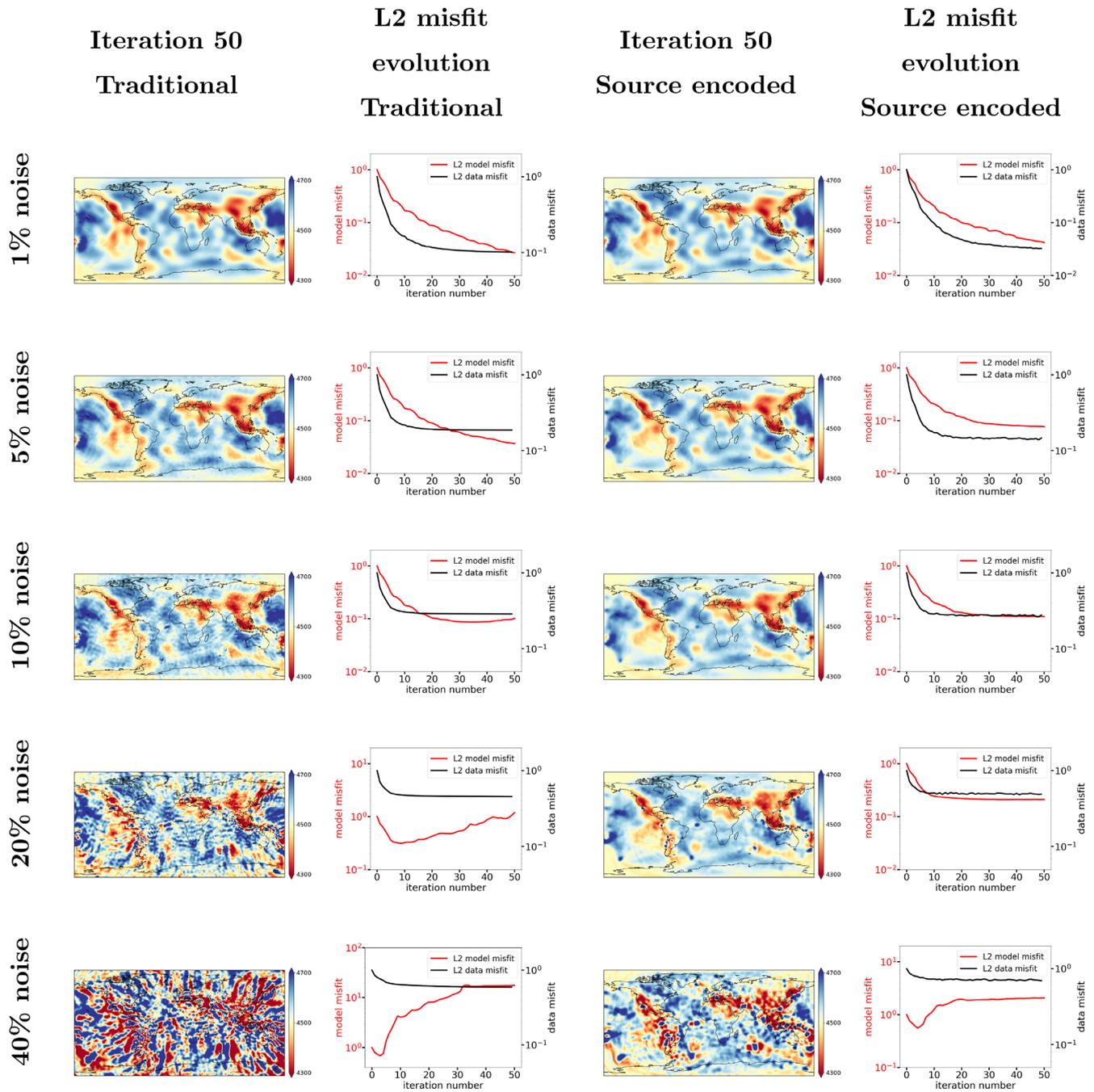


Figure 12. Traditional (first and second columns) and source-encoded (third and fourth columns) inversions when random noise at levels of 1, 5, 10, 20 and 40 per cent (rows 1–5) is added to the data. Unlike traditional FWI, source-encoded FWI remains stable when more than 10 per cent of random noise is added to the data.

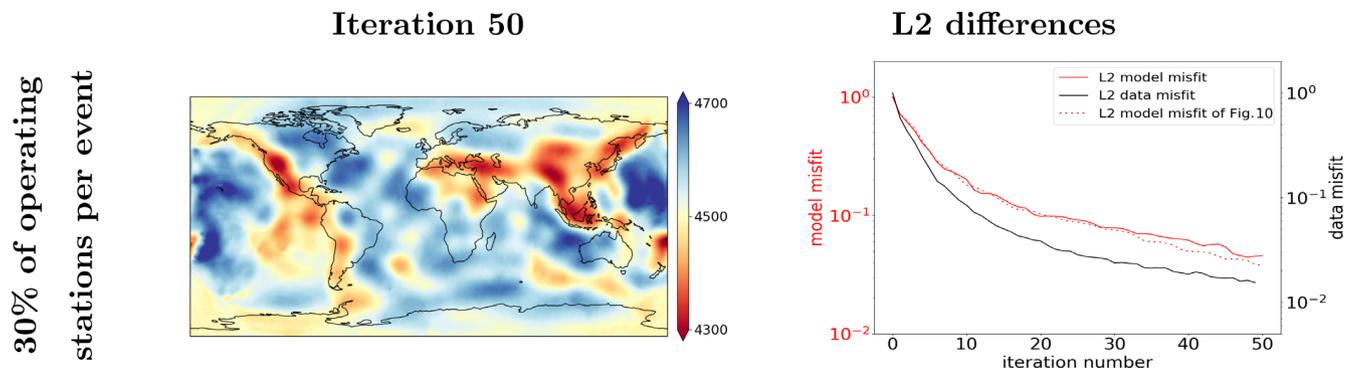


Figure 13. Same set-up as Fig. 10, but using only 30 per cent of the 293 available seismographic stations. For each event, the selected station subset is different, in order to mimic real use cases. To perform such inversions, a mask must be applied when computing the adjoint source and the misfit function, but no major changes are required.

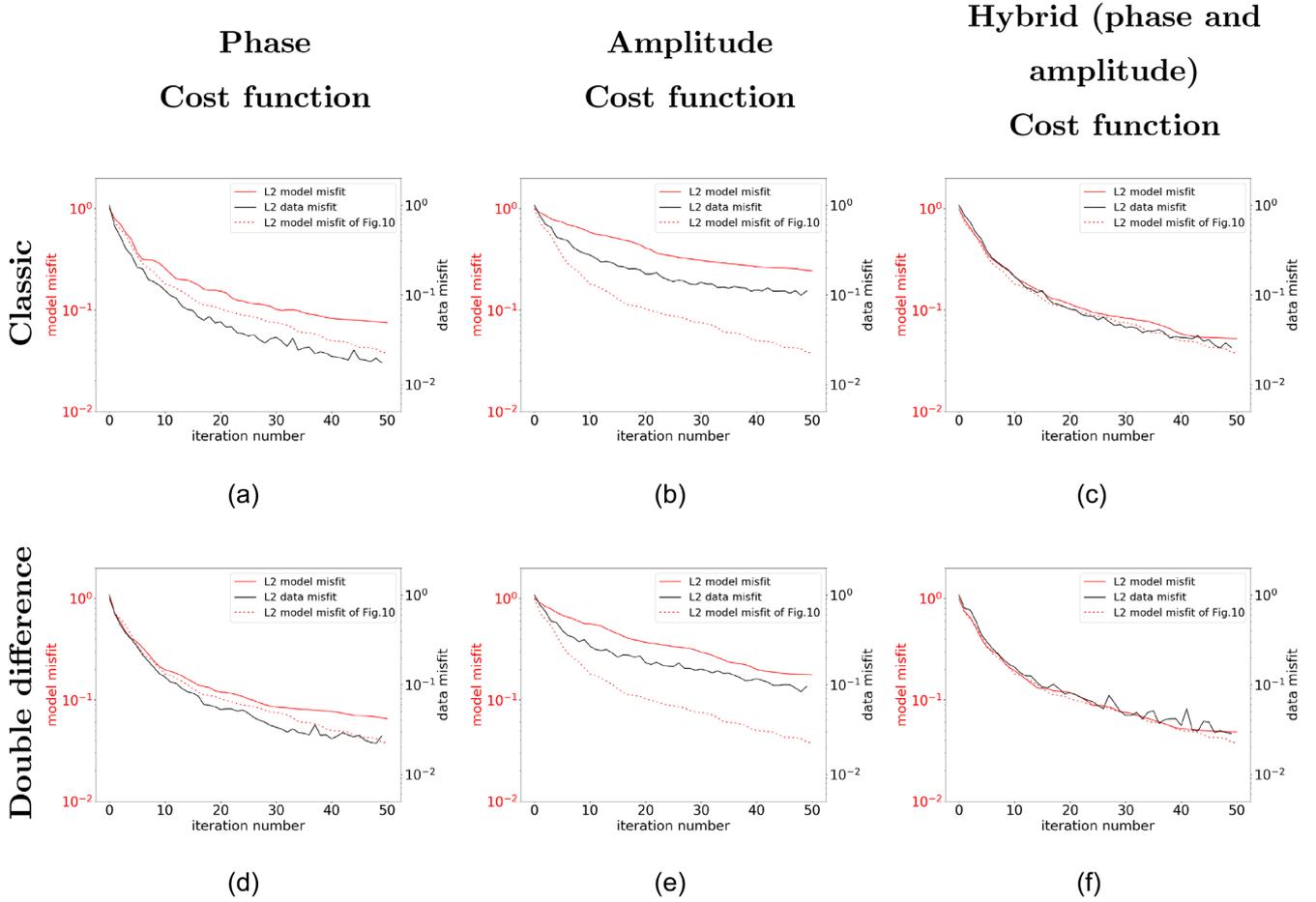


Figure 14. Comparison of the effects of different cost functions to assess the impact on convergence, using the same set-up as Fig. 10. (a) Phase only. (b) Amplitude only. (c) Hybrid. (d) Double-difference phase only. (e) Double-difference amplitude only. (f) Double-difference hybrid. Model space convergence for the waveform cost function (Fig. 10) is shown as a reference.

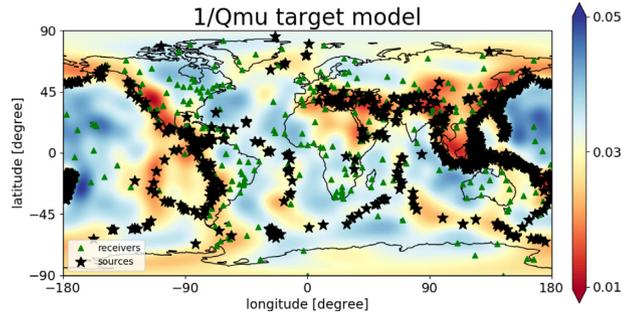


Figure 15. Target $1/Q_\mu$ model used to illustrate a shear attenuation inversion. For this problem, 1024 events labelled by black stars are used to increase data coverage, which helps to deal with the low data sensitivity to slight local $1/Q_\mu$ variations.

Since no forward wavefield storage is required for making measurements, I/O remains basically unchanged, and thus

$$I/O_{\text{encoded+measurements}} \simeq I/O_{\text{encoded}}, \quad (64)$$

even if the relative I/O cost due to writing individual seismograms may become significant for a large number of events.

Thus, given the duration of a seismogram for making measurements T , the time to reach steady state T_{ss} , and a simulation bandwidth $f_{\max} - f_{\min}$, eqs (56)–(60) and eqs (61)–(64) enable us to quantitatively assess the computational benefits of source encoding, as illustrated in Fig. 1.

A case of particular interest involves global FWI. Such computationally demanding tomographic inversions are currently carried out in the 9–150 s period range using databases with hundreds to thousands of earthquakes and three-component seismograms 60–180 min in duration (Bozdağ *et al.* 2016; Lei *et al.* 2019). Parsimonious storage requirements for kernel calculations involve more than a petabyte of disk space. Fig. 2 illustrates that if traditional measurements are made, the best we can do is a 3X speed-up in simulation time, as predicted by

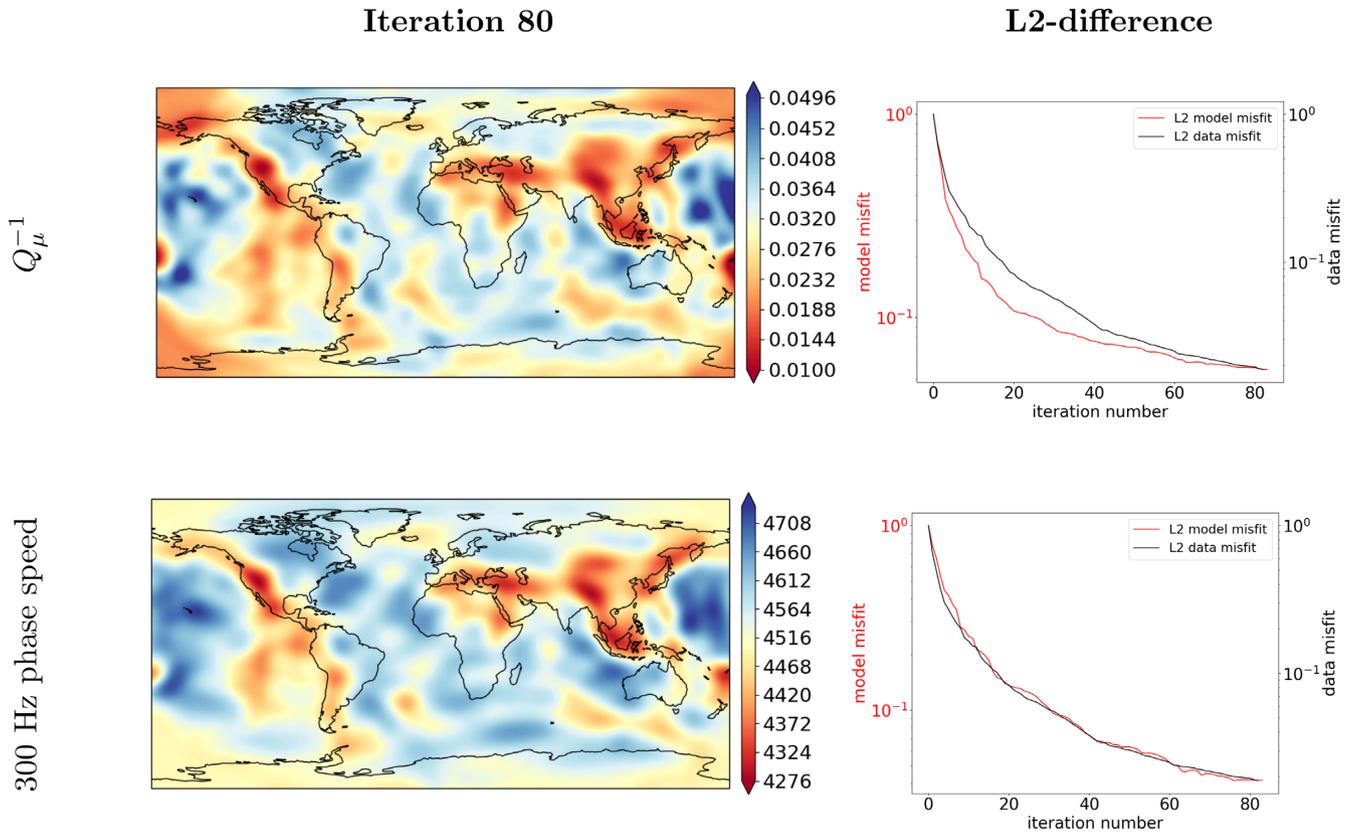


Figure 16. Illustration of a joint attenuation-wave speed inversion using source encoding. At each iteration, one frequency of each of the 1024 events is encoded, using a [200;600] Hz bandwidth. Each super wave simulation requires 53 200 time steps Δt (such that $T_{ss}/\Delta t = 2000$ and $\Delta \tau/\Delta t = 51\,200$), while a conventional simulation uses 2000 time steps. This results in a speed up of ~ 38 compared to the usual case, with a similar decrease in I/O.

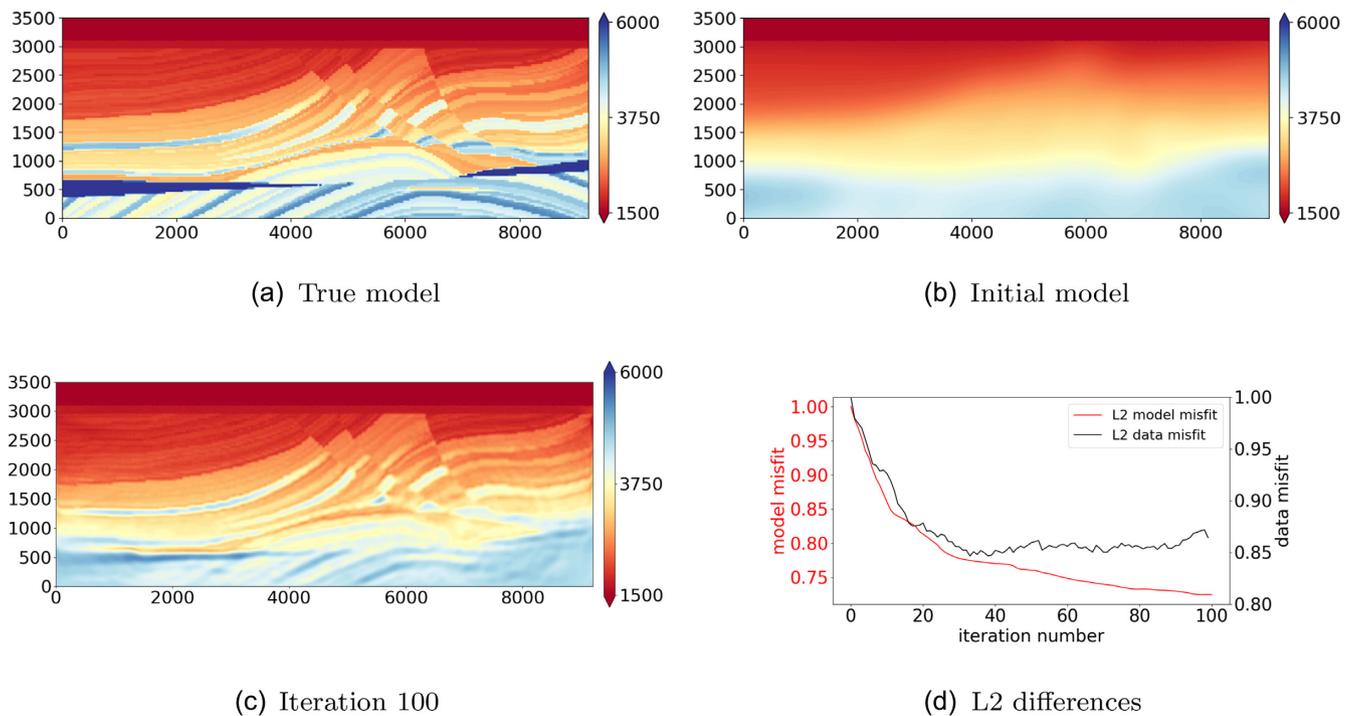


Figure 17. Marmousi offshore acoustic model inverted using source encoding. The experiment mimics a marine streamer data acquisition in which 460 sources are uniformly placed 10 m below the water surface. Each shot is recorded by a uniform array of 600 receivers located to the left of each source. The array is 6 km long, or shorter when the source is located less than 6 km from the left edge. The [1;5.6] Hz bandwidth is simultaneously inverted, using one randomized frequency per event at each iteration and a double-difference phase cost function. The frequency band is increased by 0.1 Hz at each iteration, which explains the global increase of the evolving data misfit during the second half of the inversion. Each super wave simulation requires 120 000 time steps, while a non encoded simulation requires 20 000 time steps.

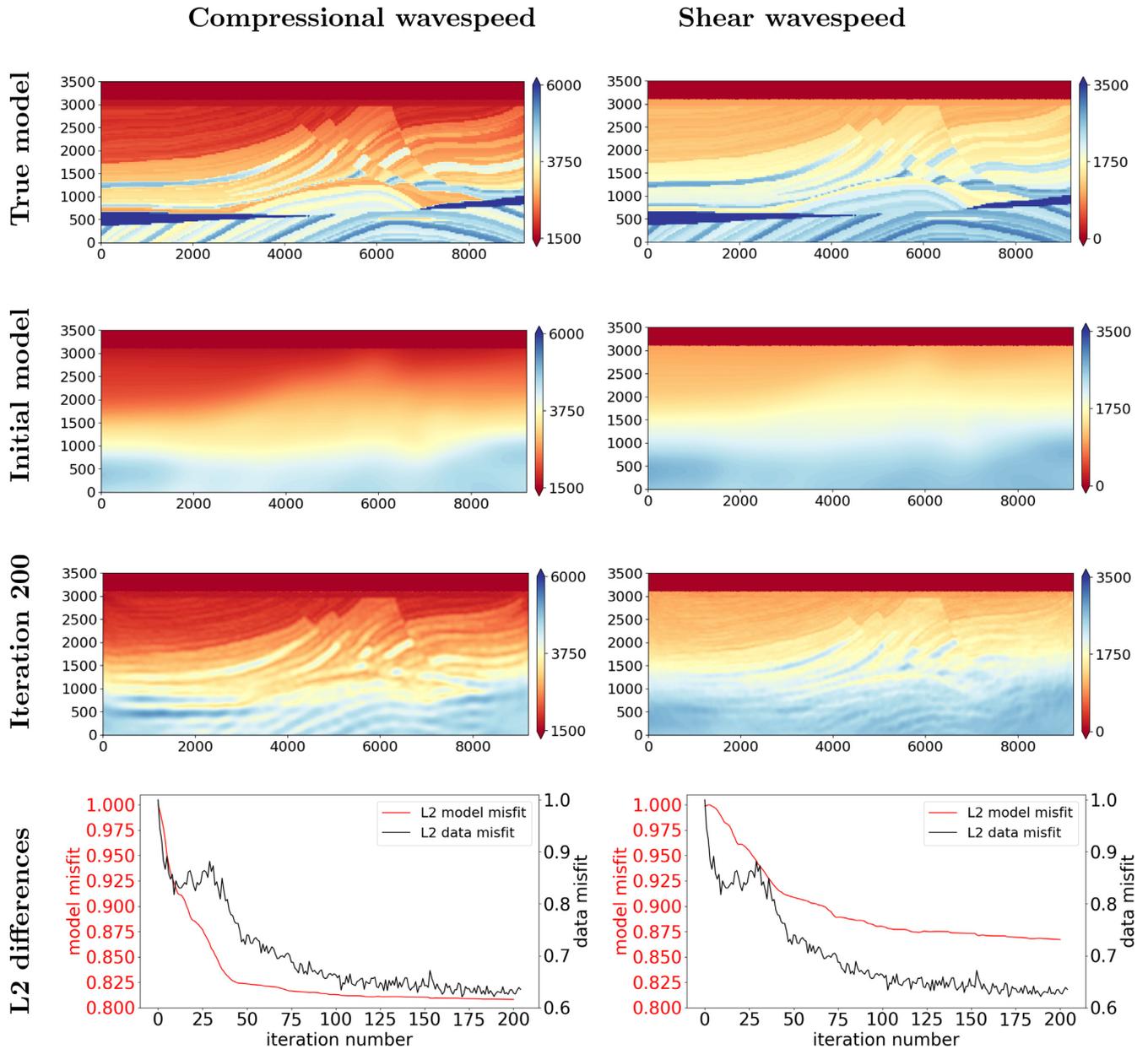


Figure 18. Coupled acoustic-elastic offshore example based on an elastic version of the Marmousi model. The [1:9] Hz bandwidth is inverted from low to high frequencies, with two frequencies assigned per source, randomized over iterations, using a double-difference phase cost function. The first column shows the results for the compressional wave speed, and the second column shows the results for the shear wave speed. The last row shows the L2 model and data misfit. The data misfit is the same because both wave speeds are simultaneously inverted, whereas the model misfit is specific to each parameter (compressional or shear wave speed).

eq. (60). The difference is much more significant for I/O, where we observe a 700X reduction in storage requirements. If global FWI could be performed without making individual measurements, which is not a simple proposition, the speed-up would be enormous, approaching factors of 700X.

14 INVERSION OPTIONS AND STRATEGIES

In this section, we illustrate the method and various inversion options and strategies. Following Yuan *et al.* (2016), we use a global Rayleigh-wave phase speed map determined by Trampert & Woodhouse (2003) as our 2-D target model. Designed for didactic purposes, this simplified model is computed on a $180 \text{ m} \times 90 \text{ m}$ rectangular mesh, with Stacey absorbing boundaries, using the membrane wave propagation solver capability of the SPECFEM2D package (Komatitsch & Tromp 1999). In the experiments, we use a combination of 32 earthquakes and 293 seismographic stations, as illustrated in Fig. 3. The ‘observed’ data were simulated using a Ricker wavelet with a 260 Hz central frequency for each event. Each wave simulation ran on a 80×40 spectral-element mesh divided into 8 MPI slices, totalling 52 552 integration

points, and using an Intel Xeon CPU E5607 cadenced at 2.27 GHz, and running computations in single precision. To perform the inversions, we used the SeisFlows framework (Modrak *et al.* 2018) and a constant phase speed map of 4500 m s^{-1} as the initial model.

14.1 Frequency-domain encoding

The first experiment is designed to demonstrate the absence of cross talk between encoded sources in the frequency domain. Fig. 4(a) shows the result of conventional FWI, described in Section 6.1, in which sensitivity kernels or ‘event kernels’ are calculated in the frequency domain for each of the 32 sources and summed to obtain the gradient of the misfit function, requiring 64 numerical simulations per iteration. The result of the source-encoded approach, described in Section 6.2, is illustrated in Fig. 4(b). In this case the gradient is calculated in the frequency domain based on eq. (50) using one super forward and one super adjoint simulation, that is, just two numerical simulations per iteration. Fig. 4(c) shows that the maximum discrepancy over space of the absolute difference between the two approaches is less than 0.0058 per cent.

14.2 Time-domain encoding

In Fig. 5 we perform the same experiment in the time domain. In this case the gradient is calculated in the time domain based on eq. (48). The small differences shown in 5(c) demonstrate that frequency-domain decoding is not necessary for the calculation of super sensitivity kernels, thanks to the mathematical equivalence of eqs (48) and (50).

14.3 Decoding

Fig. 6 compares contributions of individual events to the overall gradient calculated individually and simultaneously, and decoded based on eqs (39) and (40). Each of the 32 events is encoded with 512 frequencies, for a total of 16 384 frequencies over the [200;400] Hz bandwidth, using a frequency spacing of 0.012207 Hz. Note that despite the fact that events 1 (200.0 Hz) and 2 (200.012 Hz) are extremely close in frequency, they are still perfectly decoded without crosstalk. The absolute differences between the individual simulations (first column) and decoded super simulations (second column) are tiny (third column), further demonstrating the absence of cross talk between sources during simultaneous runs. The success of using such massive amounts of encoded frequencies also relies on the accuracy of the wave solver.

Fig. 7 summarizes investigations of the effects of (1) using an incorrect time interval for decoding and (2) decoding a time-series that has not yet reached steady state. In both cases orthogonality is lost, resulting in an incorrectly decoded signal. Note that it takes 2000 time steps to reach steady state (defining T_{ss}) followed by 1 638 400 time steps of steady state (defining $\Delta\tau$), illustrating the fact that, in general, $T_{ss} \neq \Delta\tau$, as discussed previously in Section 3.

14.4 Comparisons between classical and encoded FWI

Fig. 8 compares a usual full spectrum inversion with a source encoded inversion. The encoded version uses 32 frequencies, one per event, in the 200–400 Hz frequency range. Due to the dramatic reduction in frequency when using source encoding, we observe some oscillations in the 50th iteration source-encoded model, something we seek to mitigate in the next few experiments.

14.5 Randomization

Fig. 9 is a repeat of the previous test, but using randomization of the source frequencies used for each iteration. We conclude that randomly changing the source frequencies eliminates model oscillations, and we observe an increased convergence rate compared to the previous experiment.

14.6 Multiple frequencies per source

Another option is to assign more than one frequency per event. In this case the time interval required for decoding is given by eq. (11). The relative reduction in computational cost, given by eq. (58), becomes

$$\frac{T_{\text{traditional}}}{T_{\text{encoded}}} \simeq T(f_{\text{max}} - f_{\text{min}})/N_f, \quad (65)$$

that is, the gain is reduced by a factor N_f . Fig. 10 summarizes the results of such a strategy, using a wider 150–750 Hz frequency range to accommodate three frequencies per source. Again we observe an increased convergence rate compared to the experiment summarized in Fig. 8.

14.7 Effects of noise

Next, we explore the effects of noise on regular and source encoded inversions. A Gaussian perturbation in phase is added to each frequency in the [50;750] Hz range, with zero mean and a standard deviation of $2\pi i$, where i denotes a percentage perturbation. The effects of such noise on seismograms are illustrated in Fig. 11.

Fig. 12 shows the results for traditional and source encoded inversions when random noise at levels of 1, 5, 10, 20 and 40 per cent is added to the data. Unlike traditional FWI, we see that source-encoded inversions remain stable when more than 10 per cent of random noise is added to the data, and both the L2 model and data misfits are reduced in a stable fashion.

14.8 Variable number of receivers

A key benefit of our approach to source encoding is that every source need not be recorded by every receiver. This is basically always the case in real inversions, because some stations may not record a given event, or the data quality of certain receivers may be considered too poor for use in the inversion. To accommodate this situation, a mask must be applied when computing the super adjoint source and the misfit function, but no major changes are required. Fig. 13 demonstrates excellent convergence even when each source is recorded by a different random subset of only 30 per cent of the 293 available seismographic stations. The difference in terms of convergence rate relative to an inversion based on the full data set (Fig. 10) is minimal.

14.9 Phase, amplitude, and double-difference measurements

Next, we consider source-encoded inversions based on phase or amplitude measurements, as described in Section B. Fig. 14(a) shows that convergence in model space is slightly slower than source-encoded waveform inversion when using phase measurements (see Section B1), and Fig. 14(b) shows that convergence is even slower when using amplitude measurements (see Section B2). However, Fig. 14(c) shows that using a hybrid approach (see Section B3), combining phase and amplitude measurements, leads to a similar convergence rate as the source-encoded FWI inversion shown in Fig. 10, which used a waveform misfit.

As discussed by Yuan *et al.* (2016), double-difference measurements help reduce inversion problems related to source uncertainties. Fig. 14(d) shows the results of a double-difference phase inversion, as described in Section B4, and Fig. 14(e) shows the results of a double-difference amplitude inversion, as described in Section B5. For each event, we take double differences between all station pairs in the database. A hybrid double-difference inversion is shown in Fig. 14(f), as described in Section B6. We note that the double-difference cost functions behave similar to their classical counterparts in terms of L2 data and model misfit reduction as a function of iteration number.

14.10 Attenuation

Next, we consider a synthetic model that includes the effects of attenuation. The wave speed map, with a reference frequency of 300 Hz, is the same as in Fig. 3. The target attenuation map and the source receiver configuration are shown in Fig. 15. To both illustrate the source encoding efficiency and to better constrain the underlying optimization problem, we consider 1024 events. The joint inversion result based on a double difference hybrid cost function is shown in Fig. 16. We use the time domain sensitivity kernel computation with forward wavefield backward reconstruction, which is made possible in the presence of attenuation with an implementation of the strategy described in Komatitsch *et al.* (2016). Thanks to the improved coverage, we observe that the L2 model misfit of the wave speed decreases more than in the pure wave speed inversion. Global convergence is also obtained for the Q_{mu}^{-1} map.

14.11 Offshore acoustic FWI

To illustrate an application in exploration seismology, Fig. 17 shows the results of a source-encoded inversion for the Marmousi model created in 1988 by the Institut Français du Pétrole (see, e.g. Versteeg 2001), which is overlain by a 350 m water layer. Mimicking a streamer-type acquisition, there are 460 acoustic sources recorded by 600 hydrophones placed to the left of each source, spaced uniformly 10 m below the top of the model. These hydrophones cover 6 km, or less when the source is located less than 6 km from the left side of the model. Unlike in the preceding surface-wave examples, the challenge in this case is recovery of the model from one-sided illumination and acquisition. Each source is encoded with one randomized frequency in the 1–5.6 Hz range, and each super shot involves 120 000 time steps. The resulting speed-up compared to the 20 000 time steps required for a regular simulation is around 77. Fig. 17(c) shows the resulting model after 400 iterations based on a double-difference phase cost function. As the frequency content of the inverted data increases with each iteration, we notice an increase of the evolving data misfit during the second half of the inversion.

14.12 Offshore coupled acoustic-elastic FWI

As a second exploration seismology example, Fig. 18 shows the results of a source-encoded inversion for a coupled acoustic-elastic offshore version of the Marmousi model. The elastic model is overlain by a 350 m water layer, as shown in Fig. 17(a). There are 150 acoustic sources

recorded by 500 hydrophones placed 10 m below the top of the water layer, providing one-sided illumination and acquisition. Each source is encoded with two randomized frequencies, using a double-difference phase cost function. The first iteration uses waves in the 1–6 Hz range. At each subsequent iteration, the range is increased by 0.1 Hz until iteration 30. Iterations above 30 are performed in the 4–9 Hz range. Each super shot involves 50 000 time steps, while a non encoded simulation requires 10 000 time steps, leading to a speed-up of 30. Fig. 17(c) shows the resulting compressional and shear wave speed models after 200 iterations. Despite the fact that both the sources and receivers are located near the top of the water layer, the shear wave speed model is reasonably well recovered. Adding a few ocean bottom seismometers would undoubtedly improve the quality of the recovered shear wave speed model, but this is not the objective of this study.

15 CONCLUSION

We have further developed a version of source encoding for explicit time-domain solvers that facilitates the calculation of the gradient of a misfit function independent of the number of sources or receivers. The approach does not suffer from crosstalk between different sources, and all receivers do not need to record all sources. To obtain the gradient, source-encoded forward and adjoint wavefields are run until they reach steady state. At that point the steady-state fields are either decoded to extract their stationary parts, which are subsequently combined to obtain the full gradient by summing their respective contributions. Alternatively, the steady-state forward and adjoint wavefields are convolved over a time period proportional to the inverse of the encoded frequency spacing, which mathematically results in the same gradient. The second approach involves only minor modifications of the currently used adjoint tomography workflow.

In adjoint tomography, one generally compares individual observed and simulated seismograms by making measurements in specific time windows. In this approach, the number of numerical simulations required for making measurements and constructing adjoint sources scales linearly with the number of seismic sources, S . Once the synthetics have been calculated, the gradient of the misfit function may be calculated based on source encoding in just two additional numerical simulations per iteration, which means that one iteration requires $S + 2$ simulations. The overall computational cost may be reduced to just two simulations per iteration if the measurements needed for the construction of the adjoint sources are based on the source-encoded Fourier coefficients of the observed and simulated seismograms, leading to a single ‘super measurement’. We have demonstrated that the latter approach is stable for acoustic and offshore coupled acoustic-elastic FWI.

ACKNOWLEDGEMENTS

We thank Nicolas Boumal, Weinan E and Hadrien Montanelli for advice regarding trigonometric interpolation based on nonuniform fast Fourier transformation. We also thank reviewers Anatoly Baumstein and David Al-Attar and editors Rene-Edouard Plessix and Jean Virieux for comments and suggestions which helped to improve an earlier version of the manuscript. This research used resources provided by the Princeton Institute for Computational Science & Engineering (PICSciE), and was supported by NSF grant 1644826.

REFERENCES

- Aki, K. & Richards, P.G., 1980. *Quantitative Seismology, Theory and Methods*, W. H. Freeman, pp. 700.
- Anderson, J., Tan, L. & Wang, D., 2012. Time-reversal checkpointing methods for RTM and FWI, *Geophysics*, **77**, S93–S103.
- Ben-Hadj-Ali, H., Operto, S. & Virieux, J., 2009. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources, *Geophysics*, **76**, R109–R124.
- Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.*, **185**(2), 845–870.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.*, **207**(3), 1739–1766.
- Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.*, **162**, 541–554.
- Castellanos, C., Métivier, L., Operto, S., Brossier, R. & Virieux, J., 2015. Fast full waveform inversion with source encoding and second-order optimization methods, *Geophys. J. Int.*, **200**, 718–742.
- Chen, M., Niu, F., Liu, Q., Tromp, J. & Zheng, X., 2015. Multi parameter adjoint tomography of the crust and upper mantle beneath East Asia: 1. Model construction and comparisons, *J. geophys. Res.*, **120**, 1762–1786.
- Choi, Y. & Alkhalifah, T., 2011. Source-independent time-domain waveform inversion using convolved wavefields, *Geophysics*, **76**, R125–R134.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton Univ. Press.
- Dai, W., Huang, Y. & Schuster, G., 2013. Least-squares reverse time migration of marine data with frequency-selection encoding, *Geophysics*, **78**, S233–S242.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**(3), 1703–1725.
- Huang, Y. & Schuster, G., 2012. Multisource least-squares migration of marine streamer and land data with frequency-division encoding, *Geophys. Prospect.*, **60**, 663–680.
- Huang, Y. & Schuster, G., 2013. Multisource full waveform inversion of marine streamer data with frequency selection, *EAGE*, **66**(7), 1243–1257.
- Huang, Y. & Schuster, G., 2018. Full-waveform inversion with multisource frequency selection of marine streamer data, *Geophys. Prospect.*, **66**, 1243–1257.
- Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.*, **139**(3), 806–822.
- Komatitsch, D., Ritsema, J. & Tromp, J., 2002. The spectral-element method, Beowulf computing, and global seismology, *Science*, **298**(5599), 1737–1742.
- Komatitsch, D., Xie, Z., Bozdağ, E., Sales de Andrade, E., Peter, D., Liu, Q. & Tromp, J., 2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion, *Geophys. J. Int.*, **206**(3), 1467–1478.
- Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. & Lacasse, M., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics*, **74**, WCC177–WCC188.

- Krebs, J.R., Cha, Y., Lee, S., Dimitrov, P., Mullur, A., Downey, N. & Routh, P., 2013. Orthogonal source and receiver encoding, US Patent Application (Patent # 10,012,745).
- Lei, W. *et al.*, 2019. Global adjoint tomography – Model GLAD-M25, *Geophys. J. Int.*, submitted.
- Luo, Y., Zhu, H., Nissen-Meyer, T., Morency, C. & Tromp, J., 2009. Seismic modeling and imaging based upon spectral-element and adjoint methods, *Leading Edge*, 568–574.
- Luo, Y., Tromp, J., Denel, B. & Calandra, H., 2013. 3D coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods, *Geophysics*, **78**(4), S193–S202.
- Marquering, H., Nolet, G. & Dahlen, F., 1998. Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.*, **132**, 521–534.
- Modrak, R.T., Borisov, D., Lefebvre, M. & Tromp, J., 2018. Seisflows – flexible waveform inversion software, *Comput. Geosci.*, **115**, 88–95.
- Operto, S., Brossier, R., Combe, L., Métivier, L., Ribodetti, A. & Virieux, J., 2014. Computationally-efficient three-dimensional acoustic finite-difference frequency-domain seismic modeling in vertical transversely isotropic media with sparse direct solver, *Geophysics*, **79**(5), T257–T275.
- Örsvuran, R., Bozdağ, E., Modrak, R., Lei, W. & Ruan, Y., 2019. Towards double-difference global adjoint tomography, *Geophys. J. Int.*, submitted.
- Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**(2), 495–503.
- Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: theory and verification in a physical scale model, *Geophysics*, **64**, 888–901.
- Pratt, R.G. & Shipp, R.M., 1999. Seismic waveform inversion in the frequency domain, Part 2: fault delineation in sediments using crosshole data, *Geophysics*, **64**, 902–914.
- Pratt, R.G., Shin, C.S. & Hicks, G.J., 1998. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion, *Geophys. J. Int.*, **133**, 341–362.
- Romero, L., Ghiglia, D., Ober, C. & Morton, S., 2000. Phase encoding of shot records in prestack migration, *Geophysics*, **65**, 426–436.
- Ruan, Y., Lei, W., Modrak, R., Örsvuran, R., Bozdağ, E. & Tromp, J., 2019. Balancing unevenly distributed data in seismic tomography: a global adjoint tomography example, *Geophys. J. Int.*, submitted.
- Schiemanz, A. & Igel, H., 2013. Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: application to Valhall ocean-bottom cable data, *Geophys. J. Int.*, **195**, 1970–1988.
- Schuster, G., Wang, X., Huang, Y., Dai, W. & Boonyasiriwat, C., 2011. Theory of multisource crosstalk reduction by phase-encoded statics, *Geophys. J. Int.*, **184**, 1289–1303.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**, 988–992.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**, 433–462.
- Tarantola, A., 1984a. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.
- Tarantola, A., 1984b. Linearized inversion of seismic reflection data, *Geophys. Prosp.*, **32**, 998–1015.
- Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, **51**, 1893–1903.
- Trampert, J. & Woodhouse, J.H., 2003. Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s, *Geophys. J. Int.*, **154**(1), 154–165.
- Tromp, J., Tape, C. & Liu, Q.Y., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Versteeg, R., 2001. The Marmousi experience: velocity model determination on a synthetic complex data set, *Leading Edge*, **189**, 927–936.
- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Wright, G., Javed, M., Montanelli, H. & Trefethen, L., 2015. Extension of Chebfun to periodic functions, *J. Sci. Comput.*, **37**, C554–C573.
- Yuan, Y.O., Simons, F.J. & Tromp, J., 2016. Double-difference adjoint seismic tomography, *Geophys. J. Int.*, **206**(3), 1599–1618.
- Zhang, Q., Mao, W., Zhou, H., Zhang, H. & Chen, Y., 2018. Hybrid-domain simultaneous-source full waveform inversion without crosstalk noise, *Geophys. J. Int.*, **215**, 1659–1681.
- Zhao, Z., Sen, M. & Stoffa, P., 2016. Double-plane-wave reverse time migration in the frequency domain, *Geophysics*, **81**(5), S367–S382.
- Zhu, H. & Tromp, J., 2013. Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic ocean, *Science*, **341**(6148), 871–875.
- Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.*, **5**(7), 493–498.

APPENDIX A: ENCODED FORWARD SOURCE

A1 Earthquake seismology

In spectral-element simulations of earthquakes, one typically uses a Gaussian moment-rate source time function of the form

$$\dot{S}(t) = \frac{\alpha}{\sqrt{\pi}} \exp(-\alpha^2 t^2), \quad (\text{A1})$$

such that

$$\int_{-\infty}^{\infty} \dot{S}(t) dt = 1. \quad (\text{A2})$$

The parameter α is chosen to capture the desired half duration of the source. The source–time function is

$$S(t) = \frac{1}{2} [1 + \text{erf}(t)], \quad (\text{A3})$$

where

$$\text{erf}(t) = \frac{2\alpha}{\sqrt{\pi}} \int_0^t \exp(-\alpha^2 t'^2) dt' \quad (\text{A4})$$

denotes the error function. Spectral-element simulations are typically driven by a smooth source–time function of the form (A3).

In the frequency domain we have

$$\dot{S}(\omega) = \exp(-\omega^2/4\alpha^2), \quad (\text{A5})$$

and

$$S(\omega) = (i\omega)^{-1} \exp(-\omega^2/4\alpha^2). \quad (\text{A6})$$

The resulting encoded forward adjoint source (15) takes the form

$$F_j(\mathbf{x}, t) = - \sum_{s=1}^{N_s} M_{kj}(\mathbf{x}_s) \nabla_k \delta(\mathbf{x} - \mathbf{x}_s) (\omega_s)^{-1} \exp(-\omega_s^2/4\alpha^2) \sin(\omega_s t). \quad (\text{A7})$$

A2 Exploration seismology

In exploration seismology, one sometimes uses a source time function defined by a Ricker wavelet, namely,

$$S(t) = (1 - 2\alpha^2 t^2) \exp(-\alpha^2 t^2), \quad (\text{A8})$$

which is the second derivative of a Gaussian. In the frequency domain we have

$$S(\omega) = \frac{\sqrt{\pi}}{2\alpha^3} \omega^2 \exp(-\omega^2/4\alpha^2). \quad (\text{A9})$$

The resulting encoded forward adjoint source (15) takes the form

$$F_j(\mathbf{x}, t) = \sum_{s=1}^{N_s} f_j(\mathbf{x}_s) \delta(\mathbf{x} - \mathbf{x}_s) \frac{\sqrt{\pi}}{2\alpha^3} \omega_s^2 \exp(-\omega_s^2/4\alpha^2) \cos(\omega_s t). \quad (\text{A10})$$

APPENDIX B: SOURCE-ENCODED PHASE AND AMPLITUDE MEASUREMENTS

Suppose instead of waveform differences we wish to minimize frequency-dependent phase or amplitude differences. For convenience and to avoid clutter, we omit the dependence on the component, i , on which the measurement is made. We define

$$s^s(\mathbf{x}_r) = A^s(\mathbf{x}_r) \exp[i\theta^s(\mathbf{x}_r)], \quad d^s(\mathbf{x}_r) = A_{\text{obs}}^s(\mathbf{x}_r) \exp[i\theta_{\text{obs}}^s(\mathbf{x}_r)], \quad (\text{B1})$$

where

$$\theta^s(\mathbf{x}_r) = \arctan \left[\frac{\Im s^s(\mathbf{x}_r)}{\Re s^s(\mathbf{x}_r)} \right], \quad \theta_{\text{obs}}^s(\mathbf{x}_r) = \arctan \left[\frac{\Im d^s(\mathbf{x}_r)}{\Re d^s(\mathbf{x}_r)} \right], \quad (\text{B2})$$

and

$$A^s(\mathbf{x}_r) = |s^s(\mathbf{x}_r)| = \sqrt{[\Re s^s(\mathbf{x}_r)]^2 + [\Im s^s(\mathbf{x}_r)]^2}, \quad A_{\text{obs}}^s(\mathbf{x}_r) = |d^s(\mathbf{x}_r)| = \sqrt{[\Re d^s(\mathbf{x}_r)]^2 + [\Im d^s(\mathbf{x}_r)]^2}. \quad (\text{B3})$$

In the following sections we consider various phase and amplitude measurements and their corresponding adjoint sources. We shall see that these adjoint sources just require the decoded complex Fourier coefficients $s^s(\mathbf{x}_r)$ and $d^s(\mathbf{x}_r)$, which implies that, like the waveform difference adjoint source, they may be calculated based on a single super forward calculation.

B1 Phase measurements

First, we consider the source-encoded phase misfit function

$$\chi_\theta = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} (\Delta\theta_r^s)^2. \quad (\text{B4})$$

We have defined the differential phase measurement

$$\begin{aligned} \Delta\theta_r^s &= \arctan \left\{ \frac{\Im \left[\frac{s^s(\mathbf{x}_r)}{d^s(\mathbf{x}_r)} \right]}{\Re \left[\frac{s^s(\mathbf{x}_r)}{d^s(\mathbf{x}_r)} \right]} \right\} \\ &= \theta^s(\mathbf{x}_r) - \theta_{\text{obs}}^s(\mathbf{x}_r), \end{aligned} \quad (\text{B5})$$

where the last equality holds modulo 2π . This phase misfit function has variation

$$\begin{aligned} \delta\chi_\theta &= \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta\theta_r^s \delta\theta^s(\mathbf{x}_r) \\ &= \Re \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta\theta_r^s [A^s(\mathbf{x}_r)]^{-2} [-i s^{s*}(\mathbf{x}_r)] \delta s^s(\mathbf{x}_r). \end{aligned} \quad (\text{B6})$$

The corresponding adjoint source—injected on the appropriate component i —is

$$\begin{aligned} F_{\theta}^{\dagger}(\mathbf{x}, t) &= \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta\theta_r^s [A^s(\mathbf{x}_r)]^{-2} \Re[i s^s(\mathbf{x}_r) \exp(-i\omega_s t)] \delta(\mathbf{x} - \mathbf{x}_r) \\ &= \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta\theta_r^s [A^s(\mathbf{x}_r)]^{-1} \sin[\omega_s t - \theta^s(\mathbf{x}_r)] \delta(\mathbf{x} - \mathbf{x}_r). \end{aligned} \quad (\text{B7})$$

Next, in an attempt to avoid cycle skipping problems, we consider the source-encoded exponentiated phase misfit function

$$\begin{aligned} \chi_{\theta} &= \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} |\exp[i\theta^s(\mathbf{x}_r)] - \exp[i\theta_{\text{obs}}^s(\mathbf{x}_r)]|^2 \\ &= 2 \sum_{s=1}^S \sum_{r=1}^{R_s} \sin^2\left(\frac{1}{2} \Delta\theta_r^s\right). \end{aligned} \quad (\text{B8})$$

This phase misfit function has variation

$$\begin{aligned} \delta\chi_{\theta} &= \sum_{s=1}^S \sum_{r=1}^{R_s} \sin(\Delta\theta_r^s) \delta\theta^s(\mathbf{x}_r) \\ &= \Re \sum_{s=1}^S \sum_{r=1}^{R_s} \sin(\Delta\theta_r^s) [A^s(\mathbf{x}_r)]^{-2} [-i s^{s*}(\mathbf{x}_r)] \delta s^s(\mathbf{x}_r), \end{aligned} \quad (\text{B9})$$

and the corresponding adjoint source is

$$\begin{aligned} F_{\theta}^{\dagger}(\mathbf{x}, t) &= \sum_{s=1}^S \sum_{r=1}^{R_s} \sin(\Delta\theta_r^s) [A^s(\mathbf{x}_r)]^{-2} \Re[i s^s(\mathbf{x}_r) \exp(-i\omega_s t)] \delta(\mathbf{x} - \mathbf{x}_r) \\ &= \sum_{s=1}^S \sum_{r=1}^{R_s} \sin(\Delta\theta_r^s) [A^s(\mathbf{x}_r)]^{-1} \sin[\omega_s t - \theta^s(\mathbf{x}_r)] \delta(\mathbf{x} - \mathbf{x}_r). \end{aligned} \quad (\text{B10})$$

For small values of $\Delta\theta_r^s$ this reduces to eq. (B7).

B2 Amplitude measurements

In this section, we consider the amplitude misfit function

$$\chi_A = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} (\Delta \log A_r^s)^2, \quad (\text{B11})$$

where we have defined the logarithmic amplitude measurement

$$\begin{aligned} \Delta \log A_r^s &= \log |s^s(\mathbf{x}_r)| - \log |d^s(\mathbf{x}_r)| \\ &= \log [|s^s(\mathbf{x}_r)| / |d^s(\mathbf{x}_r)|]. \end{aligned} \quad (\text{B12})$$

This amplitude misfit function has variation

$$\begin{aligned} \delta\chi_A &= \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta \log A_r^s \delta \log [A^s(\mathbf{x}_r) / A_{\text{obs}}^s(\mathbf{x}_r)] \\ &= \Re \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta \log A_r^s [A^s(\mathbf{x}_r)]^{-2} s^{s*}(\mathbf{x}_r) \delta s^s(\mathbf{x}_r). \end{aligned} \quad (\text{B13})$$

The corresponding adjoint source is

$$\begin{aligned} F_A^{\dagger}(\mathbf{x}, t) &= \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta \log A_r^s [A^s(\mathbf{x}_r)]^{-2} \Re[s^s(\mathbf{x}_r) \exp(-i\omega_s t)] \delta(\mathbf{x} - \mathbf{x}_r) \\ &= \sum_{s=1}^S \sum_{r=1}^{R_s} \Delta \log A_r^s [A^s(\mathbf{x}_r)]^{-1} \cos[\omega_s t - \theta^s(\mathbf{x}_r)] \delta(\mathbf{x} - \mathbf{x}_r). \end{aligned} \quad (\text{B14})$$

B3 Hybrid measurements

At this point, we consider the hybrid misfit function

$$\chi_H = \alpha \chi_\theta + \beta \chi_A, \quad (\text{B15})$$

where α and β are weighting coefficients which balance the contributions of the phase and amplitude measurements to the hybrid cost function. The corresponding hybrid adjoint source is

$$F_H^\dagger = \alpha F_\theta^\dagger + \beta F_A^\dagger. \quad (\text{B16})$$

B4 Double-difference phase measurements

In this section, we consider the misfit function

$$\chi_\theta^{\text{DD}} = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} [\Delta \Delta \theta_{rr'}^s]^2, \quad (\text{B17})$$

where $w_{rr'}$ is a suitably chosen weighting function (e.g. Ruan *et al.* 2019; Örsvuran *et al.* 2019), and where we have introduced the ‘double-difference’ phase measurement

$$\begin{aligned} \Delta \Delta \theta_{rr'}^s &= \arctan \left\{ \frac{\Im \left[\frac{s^s(\mathbf{x}_r) d^s(\mathbf{x}_{r'})}{d^s(\mathbf{x}_r) s^s(\mathbf{x}_{r'})} \right]}{\Re \left[\frac{s^s(\mathbf{x}_r) d^s(\mathbf{x}_{r'})}{d^s(\mathbf{x}_r) s^s(\mathbf{x}_{r'})} \right]} \right\} \\ &= \Delta \theta_r^s - \Delta \theta_{r'}^s. \end{aligned} \quad (\text{B18})$$

The last equality holds modulo 2π . The variation of (B17) is

$$\begin{aligned} \delta \chi_\theta^{\text{DD}} &= \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} \Delta \Delta \theta_{rr'}^s [\delta \theta^s(\mathbf{x}_r) - \delta \theta^s(\mathbf{x}_{r'})] \\ &= \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} \Delta \Delta \theta_{rr'}^s [A^s(\mathbf{x}_r)]^{-2} [-i s^{s*}(\mathbf{x}_r)] \delta s^s(\mathbf{x}_r) \\ &\quad - \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} \Delta \Delta \theta_{rr'}^s [A^s(\mathbf{x}_{r'})]^{-2} [-i s^{s*}(\mathbf{x}_{r'})] \delta s^s(\mathbf{x}_{r'}) \\ &= \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r' \neq r}^{R_s} w_{rr'} \Delta \Delta \theta_{rr'}^s [A^s(\mathbf{x}_r)]^{-2} [-i s^{s*}(\mathbf{x}_r)] \delta s^s(\mathbf{x}_r). \end{aligned} \quad (\text{B19})$$

This results in the super adjoint source

$$F_\theta^{\dagger \text{DD}}(\mathbf{x}, t) = \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r' \neq r}^{R_s} w_{rr'} \Delta \Delta \theta_{rr'}^s [A^s(\mathbf{x}_r)]^{-2} \Re [i s^s(\mathbf{x}_r) \exp(-i \omega_s t)] \delta(\mathbf{x} - \mathbf{x}_r). \quad (\text{B20})$$

For the exponentiated ‘double-difference’ phase we have instead

$$\begin{aligned} \chi_\theta^{\text{DD}} &= \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} |\exp(i \Delta \theta_{rr'}^s) - \exp(i \Delta \theta_{rr'}^s)|^2 \\ &= 2 \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} \sin^2 \left(\frac{1}{2} \Delta \Delta \theta_{rr'}^s \right). \end{aligned} \quad (\text{B21})$$

The variation of this misfit function is

$$\delta \chi_\theta^{\text{DD}} = \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} \sin(\Delta \Delta \theta_{rr'}^s) [\delta \theta^s(\mathbf{x}_r) - \delta \theta^s(\mathbf{x}_{r'})], \quad (\text{B22})$$

and this results in the super adjoint source

$$F_\theta^{\dagger \text{DD}}(\mathbf{x}, t) = \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r' \neq r}^{R_s} w_{rr'} \sin(\Delta \Delta \theta_{rr'}^s) [A^s(\mathbf{x}_r)]^{-2} \Re [i s^s(\mathbf{x}_r) \exp(-i \omega_s t)] \delta(\mathbf{x} - \mathbf{x}_r). \quad (\text{B23})$$

For small values of $\Delta \Delta \theta_{rr'}^s$, this reduces to (B20).

B5 Double-difference amplitude measurements

At this point, we consider the misfit function

$$\chi_A^{\text{DD}} = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} [\Delta \Delta \log A_{rr'}^s]^2, \quad (\text{B24})$$

where we have introduced the ‘double-difference’ amplitude measurement

$$\begin{aligned} \Delta \Delta \log A_{rr'}^s &= \Delta \log A_r^s - \Delta \log A_{r'}^s \\ &= \log \left[\frac{|s^s(\mathbf{x}_r)| |d^s(\mathbf{x}_{r'})|}{|d^s(\mathbf{x}_r)| |s^s(\mathbf{x}_{r'})|} \right]. \end{aligned} \quad (\text{B25})$$

This amplitude misfit function has variation

$$\delta \chi_A^{\text{DD}} = \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r'>r}^{R_s} w_{rr'} \Delta \Delta \log A_{rr'}^s \left\{ \delta \log [A^s(\mathbf{x}_r) / A_{\text{obs}}^s(\mathbf{x}_r)] - \delta \log [A^s(\mathbf{x}_{r'}) / A_{\text{obs}}^s(\mathbf{x}_{r'})] \right\}. \quad (\text{B26})$$

The corresponding adjoint source is

$$F_A^{\dagger \text{DD}}(\mathbf{x}, t) = \sum_{s=1}^S \sum_{r=1}^{R_s} \sum_{r' \neq r}^{R_s} w_{rr'} \Delta \Delta \log A_{rr'}^s [A^s(\mathbf{x}_r)]^{-2} \Re [s^s(\mathbf{x}_r) \exp(-i\omega_s t)] \delta(\mathbf{x} - \mathbf{x}_r). \quad (\text{B27})$$

B6 Double-difference hybrid measurements

Finally, we consider the double-difference hybrid misfit function

$$\chi_H^{\text{DD}} = \alpha \chi_\theta^{\text{DD}} + \beta \chi_A^{\text{DD}}, \quad (\text{B28})$$

where α and β are weighting coefficients which balance the contributions of the double-difference phase and amplitude measurements to the cost function. The corresponding double-difference hybrid adjoint source is

$$F_H^{\dagger \text{DD}} = \alpha F_\theta^{\dagger \text{DD}} + \beta F_A^{\dagger \text{DD}}. \quad (\text{B29})$$

APPENDIX C: CROSS-CORRELATION TRAVELTIME INVERSION

In this section, we develop a source-encoded version of classical traveltime tomography.

C1 Classical traveltime inversion

Let ΔT_{sr} denote the cross-correlation traveltime anomaly of an arrival observed at receiver r due to source s . The misfit for this event is of the form (Tromp *et al.* 2005)

$$\chi_s = \frac{1}{2} \sum_{r=1}^{R_s} \left(\frac{\Delta T_{sr}}{\sigma_{sr}} \right)^2, \quad (\text{C1})$$

where σ_{sr} denotes the standard deviation. The variation of the misfit (C1) is (Tromp *et al.* 2005)

$$\delta \chi_s = \sum_{r=1}^{R_s} \frac{\Delta T_{sr}}{\sigma_{sr}^2} \frac{1}{N_{sr}} \int \dot{S}_i^s(\mathbf{x}_r, t) \delta s_i^s(\mathbf{x}_r, t) dt, \quad (\text{C2})$$

where, for a given component i ,

$$N_{sr} = \int s_i^s(\mathbf{x}_r, t) \ddot{s}_i^s(\mathbf{x}_r, t) dt, \quad (\text{C3})$$

and the associated adjoint source is (Tromp *et al.* 2005)

$$f_j^{\dagger s}(\mathbf{x}, t) = \sum_{r=1}^{R_s} \frac{\Delta T_{sr}}{\sigma_{sr}^2} \frac{1}{N_{sr}} \dot{S}_j^s(\mathbf{x}_r, -t) \delta(\mathbf{x} - \mathbf{x}_r). \quad (\text{C4})$$

This adjoint source gives rise to an associated adjoint wavefield $s_i^\dagger(\mathbf{x}, t)$, and the associated Fréchet derivatives correspond to the classical ‘banana–doughnut’ kernels (Marquering *et al.* 1998).

C2 Source-encoded traveltine inversion

The full traveltine misfit is obtained by summing the event misfit (C1) over all contributing sources:

$$\chi = \sum_{s=1}^S \chi_s = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} \left(\frac{\Delta T_{sr}}{\sigma_{sr}} \right)^2. \quad (\text{C5})$$

Its variation is

$$\delta \chi = \sum_{s=1}^S \sum_{r=1}^{R_s} \frac{\Delta T_{sr}}{\sigma_{sr}^2} \frac{1}{N_{sr}} \int \dot{S}_i^s(\mathbf{x}_r, t) \delta s_i^s(\mathbf{x}_r, t) dt. \quad (\text{C6})$$

To obtain the source-encoded version of this, using Plancherel's theorem, we rewrite the variation of the misfit function (C6) in the frequency domain as

$$\delta \chi = -\Re \sum_{s=1}^S \sum_{r=1}^{R_s} \frac{\Delta T_{sr}}{\sigma_{sr}^2} \frac{1}{N_{sr}} \int i \omega s_i^{s*}(\mathbf{x}_r, \omega) \delta s_i^s(\mathbf{x}_r, \omega) d\omega. \quad (\text{C7})$$

Motivated by this result we define the source-encoded cross-correlation misfit function as

$$\chi = \sum_{s=1}^S \chi_s = \frac{1}{2} \sum_{s=1}^S \sum_{r=1}^{R_s} \left[\frac{\Delta T_{sr}(\omega_s)}{\sigma_{sr}} \right]^2, \quad (\text{C8})$$

with variation

$$\begin{aligned} \delta \chi &= -\Re \sum_{s=1}^S \sum_{r=1}^{R_s} \frac{\Delta T_{sr}}{\sigma_{sr}^2} \frac{1}{N_{sr}} i \omega_s s_i^{s*}(\mathbf{x}_r) \delta s_i^s(\mathbf{x}_r) \\ &= -\Re \frac{1}{\Delta \tau} \int_{T_{ss}}^{T_{ss} + \Delta \tau} \int \sum_{s=1}^S \sum_{r=1}^{R_s} \frac{\Delta T_{sr}}{\sigma_{sr}^2} \frac{1}{N_{sr}} i \omega_s s_i^{s*}(\mathbf{x}_r) \delta(\mathbf{x} - \mathbf{x}_r) \exp(-i \omega_s t) \sum_{s'=1}^S \delta s_i^{s'}(\mathbf{x}) \exp(i \omega_s t) d^3 \mathbf{x} dt, \end{aligned} \quad (\text{C9})$$

where we used the orthogonality (8) to obtain the second equality. Again using the technique explained in Bozdağ *et al.* (2011) for identifying the adjoint source that corresponds to the misfit variation (C9), we find that the encoded adjoint source is determined by

$$F_j^\dagger(\mathbf{x}, t) = \Re \sum_{s=1}^S f_j^{\dagger s}(\mathbf{x}, \omega_s) \exp(i \omega_s t), \quad (\text{C10})$$

where $f_j^{\dagger s}(\mathbf{x}, \omega)$ denotes the Fourier transform of (C4). Note that this encoded adjoint source has the exact same construct as the full waveform inversion encoded adjoint source (31). The encoded adjoint source (C10) gives rise to a corresponding encoded super adjoint wavefield $S_i^\dagger(\mathbf{x}, t)$ for cross-correlation traveltine measurements.

C3 Source-encoded double-difference traveltine inversion

For 'double-difference' traveltine anomalies the regular adjoint source consists of two parts (Yuan *et al.* 2016, eqs 18a, b), namely,

$$\begin{aligned} f_j^{\dagger s}(\mathbf{x}, t) &= \sum_{r=1}^{R_s} \sum_{r' > r} \frac{\Delta \Delta T_{rr'}^s}{N_{rr'}^s} \dot{S}_j^s(\mathbf{x}_{r'}, \Delta T_{rr'}^{\text{syn}} - t) \delta(\mathbf{x} - \mathbf{x}_r) \\ &\quad - \sum_{r'=1}^{R_s} \sum_{r < r'} \frac{\Delta \Delta T_{rr'}^s}{N_{rr'}^s} \dot{S}_j^s(\mathbf{x}_r, \Delta T_{rr'}^{\text{syn}} - t) \delta(\mathbf{x} - \mathbf{x}_{r'}), \end{aligned} \quad (\text{C11})$$

where, for a given source s and receiver pair $\{r, r'\}$,

$$\Delta \Delta T_{rr'}^s = \Delta T_{rr'}^{\text{syn}} - \Delta T_{rr'}^{\text{obs}} \quad (\text{C12})$$

denotes the difference between the synthetic and observed differential crosscorrelation traveltimes, $\Delta T_{rr'}^{\text{syn}}$ and $\Delta T_{rr'}^{\text{obs}}$, respectively. The normalization factor $N_{rr'}^s$ is defined by (Yuan *et al.* 2016, eqs 16)

$$N_{rr'}^s = \int \dot{S}_i^s(\mathbf{x}_r, t + \Delta T_{rr'}^{\text{syn}}) \dot{S}_i^s(\mathbf{x}_{r'}, t) dt. \quad (\text{C13})$$

The source-encoded adjoint source is of the form (C10), where in this case $f_j^{\dagger s}(\mathbf{x}, \omega)$ denotes the Fourier transform of (C11).