Geophysical Prospecting

EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Geophysical Prospecting, 2012, **60**, 663–680 doi: 10.1111/j.1365-2478.2012.01086.x

Multisource least-squares migration of marine streamer and land data with frequency-division encoding

Yunsong Huang* and Gerard T. Schuster

King Abdullah University of Science and Technology, 4700 King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

Received July 2011, revision accepted March 2012

ABSTRACT

Multisource migration of phase-encoded supergathers has shown great promise in reducing the computational cost of conventional migration. The accompanying crosstalk noise, in addition to the migration footprint, can be reduced by least-squares inversion. But the application of this approach to marine streamer data is hampered by the mismatch between the limited number of live traces/shot recorded in the field and the pervasive number of traces generated by the finite-difference modelling method. This leads to a strong mismatch in the misfit function and results in strong artefacts (crosstalk) in the multisource least-squares migration image. To eliminate this noise, we present a frequency-division multiplexing (FDM) strategy with iterative least-squares migration (ILSM) of supergathers. The key idea is, at each ILSM iteration, to assign a unique frequency band to each shot gather. In this case there is no overlap in the crosstalk spectrum of each migrated shot gather $\mathbf{m}(x, \omega_i)$, so the spectral crosstalk product $\mathbf{m}(x, \omega_i)\mathbf{m}(x, \omega_j) = \delta_{i,j}$ is zero, unless $i = j$. Our results in applying this method to 2D marine data for a SEG/EAGE salt model show better resolved images than standard migration computed at about $1/10^{th}$ of the cost. Similar results are achieved after applying this method to synthetic data for a 3D SEG/EAGE salt model, except the acquisition geometry is similar to that of a marine OBS survey. Here, the speedup of this method over conventional migration is more than 10. We conclude that multisource migration for a marine geometry can be successfully achieved by a frequency-division encoding strategy, as long as crosstalk-prone sources are segregated in their spectral content. This is both the strength and the potential limitation of this method.

Key words: Supergather, Crosstalk, Encoded.

INTRODUCTION

For large datasets, 3D prestack wave equation migration is a computationally expensive procedure. Its computational workload is proportional to both the number of shots in a survey and the computational complexity of solving the 3D wave equation for a given velocity model. In the case of iterative methods, such as full waveform inversion (FWI), this workload is proportional to the number of iterations for acceptable convergence.

An attempt to reduce this workload was proposed by Morton and Ober (1998) by migrating one blended supergather, rather than separately migrating individual shot gathers. Here, the supergather is computed by summing a number of shot gathers, each encoded by correlation with a distinct random time series approximately orthogonal with one another. The migration image is then formed by applying a decoded

*E-mail: yunsongh@gmail.com

migration operator whose imaging condition is tuned to decoding the simultaneous sum of the encoded shots. Applying this migration operator to the supergather produces a migration image of good quality only if the number of iterations is sufficiently large. In fact, their results did not show a clear computational cost advantage over the conventional method of wave equation migration.

To mitigate problems associated with wavelets of long random time series, Jing et al. (2000) and Krebs et al. (2009) proposed a polarity encoder that randomly multiplied shot gathers by either $+1$ or -1. For phase-encoded multi-source migration, Jing et al. (2000) empirically concluded that the crosstalk term was adequately suppressed when six shot gathers were encoded, summed together and migrated. On the other hand, Krebs et al. (2009) empirically found that using this strategy with FWI produced acceptable velocity tomograms at a cost saving of at least one order of magnitude. In one of the few exceptions, Gao, Atle and Williamson (2010) used deterministic encoding to determine a shot's scale factor that gave the most significant update to the velocity model for a specified composite source. Another form of deterministic encoding is plane-wave decomposition (see e.g., Whitmore and Garing 1993; Duquet, Lailly and Ehinger 2001; Zhang et al. 2003), which also aims at reducing data volume. Using this method, Vigh and Starr (2008) obtained speedups ranging from 3 to 10-fold. Other groups, such as Virieux and Operto (2009), Dai and Schuster (2009), Boonyasiriwat and Schuster (2010), Ben-Hadj-Ali, Operto and Virieux (2009, 2011) and Ben-Hadj-Ali et al. (2011) discovered similar cost savings for FWI or least-squares migration, except that they used somewhat different encoding recipes such as exclusive use or combinations of random time shifting, frequency selection, source selection, amplitude encoding and/or spatial randomization of source locations. A related inversion scheme is by Tang (2009), who used random phase-encoding of simultaneous sources to efficiently compute the Hessian for iterative least-squares migration. Almost all of these schemes aimed to efficiently approximate the orthogonality between different encoders in as few iterations as possible.

Is there an encoding scheme that can exactly satisfy this orthogonality condition? The answer is yes. The frequency-division multiplexing (FDM) scheme from the communications industry can be used to assign each shot gather to a unique set of frequencies. Careful assignment ensures no overlap in frequencies from one shot gather to the next, thereby eliminating the crosstalk. Just as important, FDM also mitigates the acquisition crosstalk noise associated with marine geometry.

The marine acquisition crosstalk is defined as the migration noise caused by the mismatch in the modelled traces and the recorded traces. In a marine survey the recorded traces are only alive over a moving swath of hydrophones while the generated finite-difference traces are alive everywhere. This induces large residuals in the data misfit functions, leading to large artefacts in the FWI or migration images. As will be discussed later, the FDM strategy eliminates this problem. The downside of this strategy is, however, the reduced resolving power of seismic illumination. To enhance the resolving power, we use the ILSM method (Nemeth, Wu and Schuster 1999; Duquet, Marfurt and Dellinger 2000; Tang and Biondi 2009), varying each shot gather's unique frequency fingerprint at every 3 CG updates. The resulting migration algorithm for encoded data can be more than an order of magnitude faster than conventional migration while producing nearly the same image quality.

The rest of this paper is organized as follows. The theory section presents the theory of frequency-division encoding, how it can be used to remove the crosstalk in migrating supergathers and the I/O implications for computing systems. The method section, supplemented by appendices, defines the objective function for the frequency-division multisource algorithm, discusses the implications for optimization, and derives the computational complexity. The numerical results for both the 2D and 3D SEG/EAGE salt models are then presented in the numerical results section. Here, the 2D model is used to generate synthetic data emulating a marine survey and the 3D model is used to test the viability of the proposed technique for 3D data. The final section presents a summary and discussion.

Notations used in this paper are listed in Table 1.

THEORY

We now present the spectral encoding strategy for removing crosstalk artefacts in multisource imaging. We first identify the source spectrum in the forward modelling equation and we outline a typical phase-encoded multisource procedure, before developing the proposed frequency encoding method.

In the frequency domain a seismic trace with a source at $\mathbf{x}_s$ and a receiver at $\mathbf{x}$ can be expressed (Stolt and Benson 1986), based on the Born approximation to the Lippman-Schwinger equation, as

$$d(\mathbf{x}|\mathbf{x}_s) = \int G(\mathbf{x}|\mathbf{x}')m_o(\mathbf{x}')G(\mathbf{x}'|\mathbf{x}_s)\,W_s(\omega)\,\mathrm{d}\mathbf{x}'. \tag{1}$$

Here, $G(\mathbf{b}|\mathbf{a})$ denotes the Green's function from $\mathbf{a}$ to $\mathbf{b}$; $m_o(\mathbf{x}') \stackrel{\text{def}}{=} s(\mathbf{x}')\delta s(\mathbf{x}')$ is the reflection coefficient-like term at

Table 1 Notations used in this paper.

Notation	Description
CG	Conjugate Gradient method
FD	Frequency Division
ILSM	Iterative Least-Squares Migration
IS	Iterative Stacking (of Frequency-Division MultiSource)
j	Angular frequency index
K_{it}	Number of iterative updates
K_{CGit}	Number of iterative updates given fixed encoding
LSFDMS	Least-Squares Frequency-Division MultiSource
$\mathbf{m}$	Reflectivity model
M_{ga}	Number of supergathers
n_h	Number of receivers associated with a source
n_{htot}	Total number of receivers covered by a supergather. In marine streamer acquisition, $n_{htot} \geq n_h$.
n_ω	Number of discretized frequencies
normalized $f_{K_{it}}$	$f_{K_{it}}/f_0$, assuming $\mathbf{m}\|_{K_{it}=0} = 0$. f could be 'model error' or 'objective function'.
S	Number of sources included in a supergather
S_{tot}	Total number of sources
SNR	Signal-to-noise ratio (quantifying how the observed CSG is corrupted by noise in this paper)
	Explicit functions of ω
W	Source spectrum
$\mathbf{d}$	Data, in frequency domain
$\mathbf{L}$	prestack modelling operator
$\underline{\mathbf{L}}$	$\mathbf{L}$ deprived of W
N_s	Encoding function for source s
$\tilde{\boxtimes}$	Encoded version of $\boxtimes$

$\mathbf{x}'$, where $\delta s(\mathbf{x}')$ is the slowness perturbation from an assumed background slowness $s(\mathbf{x}')$; and $W_s(\omega)$ is the spectrum of the s^{th} source weighted by $-2\omega^2$ and can be pulled outside the integral since it is independent of $\mathbf{x}'$. For conciseness $W_s(\omega)$ is hereafter referred to as the 'source spectrum' or simply 'spectrum' for short. As the earth model is discretized into M grid points, equation (1) can be recast in matrix-vector form as

$$\mathbf{d}_s = W_s(\omega)\underline{\mathbf{L}}_s\mathbf{m}, \quad \forall s = 1, \ldots, S \tag{2}$$

which is conventionally expressed as

$$\mathbf{d}_s = \mathbf{L}_s\mathbf{m}, \quad \forall s = 1, \ldots, S \tag{3}$$

where $\mathbf{L}_s = W_s(\omega)\underline{\mathbf{L}}_s$. (4)

Here, $\mathbf{m} \in \mathbb{R}^M$ is the reflectivity model; $\mathbf{d}_s \in \mathbb{C}^{n_h}$ represents the s^{th} shot gather; S is the number of shots; n_h is the number of receivers per shot; $W_s(\omega)$ is a scalar denoting the spectrum of the s^{th} source; $\mathbf{L}_s \in \mathbb{C}^{n_h \times M}$ represents the prestack modelling operator for the s^{th} shot gather and $\underline{\mathbf{L}}_s$ is $\mathbf{L}_s$ deprived of $W_s(\omega)$. Equations (2) to (4) are in the frequency domain and recognize that quantities such as $\mathbf{d}_s$ $\underline{\mathbf{L}}_s$ and $\mathbf{L}_s$ all depend on ω, which is silent to reduce notational clutter; however, ω is explicitly retained in $W_s(\omega)$, because $W_s(\omega)$ represents the proposed frequency encoding function. Note also the subscript in $W_s(\omega)$, implying that different sources may have different spectra.

Phase encoding

Our frequency encoding scheme will now be developed in the same framework of phase-encoding (Romero et al. 2000), which typically consists in the following three steps. 1) Different shot gathers are uniquely phase encoded. 2) They are summed together to form supergathers, which are then 3) migrated all at once. The first step amounts to multiplying the s^{th} shot gather with a unique phase-encoding function N_s, a step expressed as

$$\tilde{\mathbf{d}}_s = N_s\mathbf{d}_s. \tag{5}$$

Then $\tilde{\mathbf{d}}_s$ are summed over all sources to give the encoded supergather $\tilde{\mathbf{d}}$:

$$\tilde{\mathbf{d}} = \sum_{s=1}^{S} \tilde{\mathbf{d}}_s = \sum_{s=1}^{S} N_s\mathbf{d}_s, \tag{6}$$

$$= \tilde{\mathbf{L}}\mathbf{m}, \tag{7}$$

where the multisource phase-encoded prestack modelling operator is defined as

$$\tilde{\mathbf{L}} \overset{\text{def}}{=} \sum_{s=1}^{S} N_s\mathbf{L}_s. \tag{8}$$

Finally, the third step involves applying the adjoint operator $\tilde{\mathbf{L}}^\dagger$ to the encoded supergather $\tilde{\mathbf{d}}$ in equation (6), before applying the imaging condition, to obtain the migrated image $\tilde{\mathbf{m}}$ as

$$\tilde{\mathbf{m}} = \sum_{\omega} \tilde{\mathbf{L}}^\dagger\tilde{\mathbf{d}} \tag{9}$$

$$= \sum_{\omega}\sum_{s=1}^{S}\sum_{q=1}^{S} N_s^* N_q \mathbf{L}_s^\dagger\mathbf{L}_q\mathbf{m} \tag{10}$$

$$= \hat{\mathbf{m}} + \mathbf{c}, \tag{11}$$

where

$$\hat{\mathbf{m}} \overset{\text{def}}{=} \sum_{\omega} \sum_{s=1}^{S} |N_s|^2 \mathbf{L}_s^{\dagger} \mathbf{L}_s \mathbf{m} \tag{12}$$

$$= \sum_{\omega} \sum_{s=1}^{S} \mathbf{L}_s^{\dagger} \mathbf{L}_s \mathbf{m}, \tag{13}$$

and

$$\mathbf{c} \overset{\text{def}}{=} \sum_{\omega} \sum_{s=1}^{S} \sum_{q \neq s}^{S} N_s^* N_q \mathbf{L}_s^{\dagger} \mathbf{L}_q \mathbf{m} \tag{14}$$

$$= \sum_{\omega} \sum_{s=1}^{S} \sum_{q \neq s}^{S} N_s^* N_q W_s^*(\omega) W_q(\omega) \underline{\mathbf{L}}_s^{\dagger} \underline{\mathbf{L}}_q \mathbf{m}. \tag{15}$$

Here, $\hat{\mathbf{m}}$ is the sequential shot-gather migration and $\mathbf{c}$ is crosstalk noise. Equation (13) follows assuming the phase-encoding function N_s is of pure phase so that $N_s^* N_s = 1$ and equation (15) follows from equation (4).

Note the crosstalk noise, $\mathbf{c}$, is the only part of $\tilde{\mathbf{m}}$ that depends on the random phase encoding function, over which an ensemble average, denoted by $\langle\,\rangle$, is taken to produce

$$\langle \mathbf{c} \rangle = \sum_{\omega} \sum_{s=1}^{S} \sum_{q \neq s}^{S} \langle N_s^* N_q \rangle \mathbf{L}_s^{\dagger} \mathbf{L}_q \mathbf{m}. \tag{16}$$

Frequency-division encoding

While existing approaches such as studied by Schuster et al. (2011) strive to reduce crosstalk noise by devising phase-encoding functions such that $\langle N_s^* N_q \rangle = 0$ for $s \neq q$, this paper relies on devising source spectra $W_q(\omega)$'s in order to eliminate the noise term $\mathbf{c}$ defined in equation (15). To this end, we single out an arbitrary term in equation (15) and investigate how to make it zero. An example of such a term is expressed as

$$\mathbf{c}_{sq} = \sum_{\omega} W_s^*(\omega) \boldsymbol{\nu}_{sq}(\omega) W_q(\omega), \quad s \neq q, \tag{17}$$

where

$$\boldsymbol{\nu}_{sq}(\omega) \overset{\text{def}}{=} N_s^* N_q \underline{\mathbf{L}}_s^{\dagger} \underline{\mathbf{L}}_q \mathbf{m}. \tag{18}$$

Because the dependence of $\boldsymbol{\nu}_{sq}(\omega)$ on ω is typically spatially varying and unknown, it is impossible to construct $W_q(\omega)$'s that can suppress all elements of $\mathbf{c}_{sq}$, unless the source spectra are non-overlapping. Non-overlapping source spectra ensure that

$$W_s^*(\omega) W_q(\omega) = 0, \quad \text{for} s \neq q, \ \forall \omega \tag{19}$$

and in turn reduce equation (15) to zero. We refer to this encoding scheme specifically as the frequency-division (FD) method. The previous analysis contrasts the different roles that the phase encoder N_s and frequency encoder $W_s(\omega)$ play. For notational economy, however, hereafter in the context of FD we recast N_s as a frequency encoder, on which $W_s(\omega)$ is predicated; in addition, ω is discretized, and is identified with a frequency index j running from 1 to n_ω. The frequency encoder is given as a binary vector

$$N_s(j) \overset{\text{def}}{=} \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ frequency belongs to source } s, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Note that $N_s(j)$'s are no longer of pure phase; this can be regarded as a form of amplitude encoding (Godwin and Sava 2010). If no frequency index is shared by multiple sources, then equation (20) leads to

$$N_s(j) N_q(j) = 0, \quad \text{for } s \neq q, \quad \forall j = 1, \ldots, n_\omega. \tag{21}$$

Thus equation (19) is guaranteed, in our new notation, by the choice

$$W_s(j) = N_s(j) W(j), \quad \forall s = 1, \ldots, S, \quad \forall j, \tag{22}$$

where $W(j)$ is the intact source spectrum. In shorthand, equation (21) can be expressed as

$$\mathbf{N}_s \odot \mathbf{N}_q = \mathbf{0}, \quad \text{for } s \neq q, \tag{23}$$

where $\odot$ represents element-wise multiplication between two vectors. If, moreover, every frequency index is assigned to some source, then equation (20) leads to

$$\bigoplus_{s=1}^{S} \mathbf{N}_s = \mathbf{1}, \tag{24}$$

where $\oplus$ represents element-wise addition. Accordingly, following equation (22), we have

$$\mathbf{W}_s \odot \mathbf{W}_q = \mathbf{0}, \quad \text{for } s \neq q, \tag{25}$$

$$\bigoplus_{s=1}^{S} \mathbf{W}_s = \mathbf{W}. \tag{26}$$

Given S sources and n_ω frequency indices, FD endeavours to evenly divide the latter among the former. That is, on average each source is assigned n_ω/S frequency indices.

We outline next how n_ω is determined. Suppose the maximal traveltime between sources and their associated receivers is T, the peak frequency of the source wavelet is f_0 and the cut-off high frequency is at $f_{bi} = 2.5 f_0$. The Nyquist sampling theorem dictates $dt < \frac{1}{2 f_{bi}} = \frac{1}{5 f_0}$ and therefore the total

number of time samples is $n_t = T/dt > 5Tf_0$. For real signals devoid of DC, the number of independent angular frequencies is given by

$$n_\omega = 2.5Tf_0. \tag{27}$$

For example, the parameters chosen for our 2D and 3D simulations are: $n_\omega = 160$ as $T = 2$s and $f_0 = 32$Hz and $n_\omega = 360$ as $T = 9$s and $f_0 = 16$Hz, respectively. Note that the effective number of independent frequencies is smaller than n_ω, because the source spectrum $W(j)$ is far from uniform.

Eliminating marine acquisition crosstalk

Once sources have been assigned non-overlapping sets of frequencies, marine acquisition crosstalk can be eliminated. The key idea is, after completing multisource forward modelling by computer simulation, at each receiver h any extraneous frequency component j is pruned; j is considered extraneous if j is assigned to a source, to which, according to marine geometry, h is not associated. We illustrate the proposed algorithm with an example shown in Fig. 1. Figure 1(a) depicts $S = 3$ sources, $n_\omega = 5$ frequencies and a specific way of FD described by the frequency encoders $N_s(j)$'s identified as, from left to right $N_3 = [1, 0, 0, 0, 1]^T$, $N_2 = [0, 1, 0, 0, 0]^T$ and $N_1 = [0, 0, 1, 1, 0]^T$. Figure 1(b) depicts a towed marine geometry, where each source is associated with $n_h = 5$ receivers. For instance, source s_3 is associated with receivers h_5–h_8, but not with receivers h_1–h_4. Consider for example $j = 5$ at receiver h_4. Because, according to Fig. 1(a), $j = 5$ is assigned to source s_3, $j = 5$ is considered extraneous at receivers h_4 and should be pruned. The rationale is as follows. When sources are blended, all frequency components are present (see equation (26)) in forward modelling and consequently at every receiver. Receiver h_4 would have detected frequency component 5, which comes from source s_3 but h_4 lies outside the aperture associated to s_3, and therefore h_4 should not pick up any signal stemming from s_3. This explains the pruning of the extraneous frequency component 5 at h_4. This is indicated by the absence of a bar corresponding to $j = 5$ at h_4 in Fig. 1(c). Other unoccupied frequency slots in Fig. 1(c) are likewise inferred.

The pruning operation is equivalent to selective filling in as follows. Let $\mathrm{F_{blen}}$(frequency, receiver) of size $n_\omega \times n_{htot}$ be the outcome[1] in the frequency domain detected by receivers generated by forward modelling with blended sources prior to pruning and let $\mathrm{F_{prun}}$(frequency, receiver) of the same size be the result of pruning applied to $\mathrm{F_{blen}}$. Here, n_{htot} is the total number of receivers covered by the supergather and $n_{htot} = 8$ in this example. $\mathrm{F_{prun}}$ is obtained by first initialization with 0 and subsequently filling in with valid entries in $\mathrm{F_{blen}}$; an entry $\mathrm{F_{blen}}(j, h)$ is valid if frequency component j is not extraneous at receiver h. For instance for $j = 5$, we have

$$\mathrm{F_{prun}}(j = 5, [h_5, h_6, h_7, h_8]) \leftarrow \mathrm{F_{blen}}(j = 5, [h_5, h_6, h_7, h_8]). \tag{28}$$

Similarly, the encoded supergather $\mathrm{CSG_{enc}}$, of size $n_\omega \times n_{htot}$, can be formed as follows, assuming that the observed CSG's have been transformed to the frequency domain and are indexed as CSG(frequency, receiver, source), of size

$$M_{CSG} = n_\omega \times n_h \times S. \tag{29}$$

Here, n_h is the number of receivers associated with each source in acquisition and $n_h = 4$ in this example. First, $\mathrm{CSG_{enc}} \leftarrow 0$. Next, fill in $\mathrm{CSG_{enc}}$ with the corresponding entries in CSG according to the current frequency encoders. Specifically, loop j over n_ω, and for a given j, find to which source s it belongs and subsequently find which receivers h's are associated to this s. Then execute $\mathrm{CSG_{enc}}(j, h\text{'s}) \leftarrow \mathrm{CSG}(j, :, s)$. An example for $j = 3$ is given as

$$CSG_{enc}(j = 3, [h_1, h_2, h_3, h_4]) \leftarrow CSG(j = 3, :, s_1). \tag{30}$$

Finally, the misfit function is computed by $\mathrm{F_{prun}} - \mathrm{CSG_{enc}}$. By pruning or equivalently selective filling in, the mismatch problem between the limited number, n_h, of live traces/shot in observed CSG and the pervasive number, n_{htot}, of traces in simulation-generated $\mathrm{F_{blen}}$ is now resolved.

Note that since there are n_ω equations similar to equation (30), each reading n_h entries, the total number of entries read from CSG by selective filling in is

$$M_{CSG_{enc}} = n_\omega \times n_h, \tag{31}$$

$$= \frac{M_{CSG}}{S}. \tag{32}$$

In this example $M_{CSG_{enc}} = 5 \times 4 = 20$, coinciding with the number of bars in Fig. 1(c).

Input/output implications

I/O is an important consideration when dealing with industrial-size data sets. Contrary to the naive impression that the I/O cost of the proposed method in K_{it} iterations is K_{it} times that of standard migration, here we show that the

[1] in MATLAB notation; likewise for the following arrays in this section

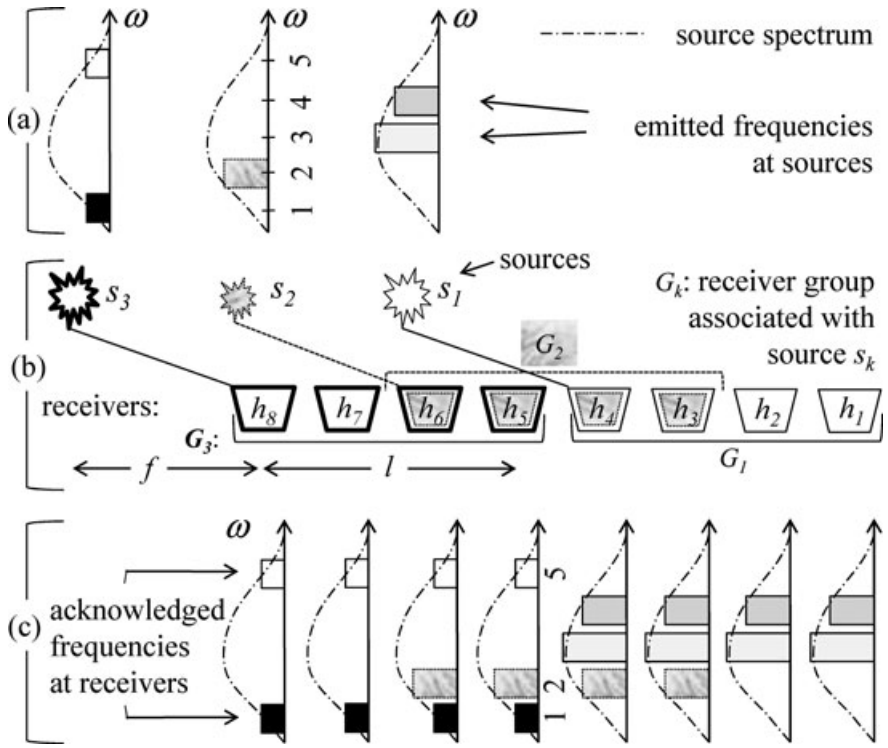

Figure 1. Frequency division of sources for one supergather of towed-marine data. Sources and receivers are identified with their indices. (a) Unique spectra assigned to and hence will be emitted by the sources. The three spectra patterns are non-overlapping. (b) The association, signified with the same line width and fill style, between sources and their respective receiver groups. f denotes near off-set; l denotes line length. (c) Frequencies listened to at each receiver.

actual I/O cost is only $2 + \epsilon$ times[2] the latter, assuming the migrations are carried out in the frequency domain while the original data are stored in the time domain.

Let the I/O cost be identified with the size[3] of data passing through I/O and assume the data size is M_{CSG}. Standard migration entails reading every shot gather, followed by Fourier transform and then the migration. So the I/O cost is $C_0 = M_{CSG}$. On the other hand, the work flow of the proposed method consists in two stages. (1) Preparation. All input data are read, transformed to the frequency domain and saved to disk. The I/O cost of this stage is $C_1 = 2M_{CSG}$. (2) Migration. The I/O cost per iteration is $M_{CSG_{enc}}$. In K_{it} iterations, the I/O cost is $C_2 = K_{it} M_{CSG_{enc}} = K_{it} \frac{M_{CSG}}{S} = \epsilon M_{CSG}$, where $\epsilon = K_{it}/S \ll 1$ as K_{it} is typically an order of magnitude smaller than S, which is how speedup can be gained by iterative multisource methods. Therefore the I/O cost $C_1 + C_2 = (2 + \epsilon)C_0$ of the proposed method is a little more than twice that of the standard approach for any $K_{it} \ll S$.

If $\mathrm{CSG}_{\mathrm{enc}}$ can fit in a computer's memory, C_2 can be further reduced as follows. Read the CSGs from disk to form a $\mathrm{CSG}_{\mathrm{enc}}$, which is kept in the memory, then make K_{CGit} iterative updates[4] to the trial model. In this scheme, C_2 is reduced by a factor[5] of K_{CGit}.

METHOD

Multisource objective function

Due to frequency division, only a subset of the spectrum will be covered at each source at each iteration and so ringy migration artefacts are expected. An effective method to reduce migration artefacts (Nemeth et al. 1999; Duquet et al. 2000) is least-squares migration (LSM), which works by iteratively updating a trial model in order to minimize a data misfit function. A widely adopted misfit function is the L_2 norm squared

[2] ϵ is a small fraction, for instance, 1/10.
[3] Measured in the number of complex numbers

[4] Since the Hessian of the objective function is a constant given a fixed $\mathrm{CSG}_{\mathrm{enc}}$, these K_{CGit} iterations are made using CG.
[5] As K_{CGit} increases, K_{it} may also have to increase in order to produce an acceptable result. Therefore this reduction factor is a bit smaller than K_{CGit}.

of data error. In addition, regularization with Cauchy norm (Amundsen 1991; Sacchi 1997; Wang and Sacchi 2007) is used in this paper. In the Bayesian framework (Aster, Borchers, and Thurber 2005; Debski 2010), the regularization corresponds to a negative logarithm of the *a priori* distribution of the model. The choice of Cauchy distribution is meant to capture the sparse nature of typical reflectivity models. Following the Bayesian approach, we write the regularization as

$$R(\mathbf{m}) = -\ln p_c(\mathbf{m}) = -\ln\left[\prod_i \frac{c}{\pi(c^2+m_i^2)}\right] \tag{33}$$

$$= \sum_i \ln\left(c^2+m_i^2\right) + \text{ constants}, \tag{34}$$

where $p_c(\mathbf{m})$ is a 0-median Cauchy distribution with parameter c; and we write the misfit function as

$$e(\mathbf{m}) = -\ln g_{\sigma^2}(\widetilde{\mathfrak{d}}|\mathbf{m}) = \frac{1}{2\sigma^2}||\widetilde{\mathfrak{d}} - \widetilde{\mathfrak{L}}\mathbf{m}||^2 + \text{ constants}, \tag{35}$$

where $g_{\sigma^2}(\cdot)$ is a 0-mean Gaussian distribution with variance σ^2. The probabilistic formulations allow us to determine the parameters c and σ^2 by maximum likelihood estimation (MLE). In equations (34) and (35) the constants are independent of $\mathbf{m}$. In equation 35, $\widetilde{\mathfrak{d}} \in \mathbb{C}^{n_{htot}n_\omega M_{ga}}$ and $\widetilde{\mathfrak{L}} \in \mathbb{C}^{n_{htot}n_\omega M_{ga}\times M}$ are respectively formed by concatenating $\tilde{\mathbf{d}}^{(\gamma,j)}$ and $\tilde{\mathbf{L}}^{(\gamma,j)}$ along the column dimension in dictionary order of (γ, j), where $\gamma = 1,\ldots, M_{ga}$ is the supergather index, with M_{ga} being the number of supergathers and $j = 1,\ldots, n_\omega$ is the frequency index. Here, the descriptor (γ, j) explicates the fact that $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{L}}$ as defined in equations (6) and (8), respectively, are specific to a particular supergather and frequency. Note that in the case of marine streamer acquisition, the first dimension of $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{L}}$ is extended from n_h to n_{htot}. In contrast, in the standard approach of a single shot gather, the counterparts of $\widetilde{\mathfrak{d}}$ and $\widetilde{\mathfrak{L}}$ would be of sizes $\mathbb{C}^{n_h n_\omega S_{tot}}$ and $\mathbb{C}^{n_h n_\omega S_{tot}\times M}$, respectively, where $S_{tot} = SM_{ga}$ is the total number of sources.

The objective function is then constructed as

$$J(\mathbf{m}) = \sigma^2(e(\mathbf{m}) + R(\mathbf{m})) = \frac{1}{2}||\widetilde{\mathfrak{d}} - \widetilde{\mathfrak{L}}\mathbf{m}||^2 + \sigma^2\sum_i \ln\left(c^2+m_i^2\right), \tag{36}$$

where additive constants have been dropped. Its negative gradient is given as

$$\mathbf{g} \stackrel{\text{def}}{=} -\nabla_{\mathbf{m}} J(\mathbf{m}) = \widetilde{\mathfrak{L}}^\dagger(\widetilde{\mathfrak{d}} - \widetilde{\mathfrak{L}}\mathbf{m}) - 2\sigma^2\sum_i Q(m_i)m_i, \tag{37}$$

where

$$Q(m_i) = \frac{1}{c^2+m_i^2}. \tag{38}$$

Note that the shape of the objective function $J(\mathbf{m})$ typically changes over iteration step k because every iteration typically requires a new pass of FD encodings for the M_{ga} supergathers to generate $\widetilde{\mathfrak{d}}$ and to effect $\widetilde{\mathfrak{L}}$. That the objective function depends on k is a topic that is studied in stochastic optimization (Spall 2003). Our problem (albeit of a much larger size) is similar to the 'stochastic bowl' studied by Schraudolph and Graepel (2002), because as shown in Appendix A the Hessian of the misfit function pertaining to FD encoded supergathers consists in terms sampled from the standard full Hessian.

As FD encoding could significantly alter the Hessian, the conjugacy condition of the Conjugate Gradient (CG) can not be maintained if supergathers are formed with new FD encoding at each iteration, a strategy known as 'dynamic encoding'. On one hand, in order to accelerate convergence, and on the other hand, in order to reduce I/O cost, we adopt a strategy of a hybrid CG (termed 'CG within mimi-batch' in Schraudolph and Graepel 2002), whereby supergathers are encoded anew every K_{CGit} iteration. $K_{CGit} = 3$ is chosen in this study. Given fixed supergathers and a fixed $Q(m_i)$ defined in equation (38), K_{CGit} iterations are carried out by a CG scheme (outlined in Algorithm 1 in Appendix C). Then supergathers are randomly encoded again; $Q(m_i)$'s are updated, which is known as the 'Iterative Reweighted Least-Squares' method (Scales, Gersztenkorn and Treitel 1988); the parameters c and σ^2 of the probability distributions are re-estimated through MLE; and the search direction of CG is reset to a negative gradient.

Migration method

The migration method is considered next. We choose prestack split-step migration based on the following two reasons. First, the fact that sources are subject to phase and/or frequency encoding demands that prestack migration is the method of choice. Second, aside from computational efficiency and the absence of operator aliasing, the fact that phase-shift migration is a spectral technique makes it particularly convenient to perform frequency encoding. To handle smooth lateral variations in the velocity field, we opt for split-step migration (Stoffa et al. 1990), as did Kuehl and Sacchi (1999). It is a straightforward procedure to adapt this to RTM, with the finite-difference method replacing the spectral method.

The use of LSM requires both the forward modelling and migration operations; and the use of prestack migration requires both a source field and downward continued data field. The details of this migration method are relegated to Appendix B, which is included because of the usefulness in assessing the

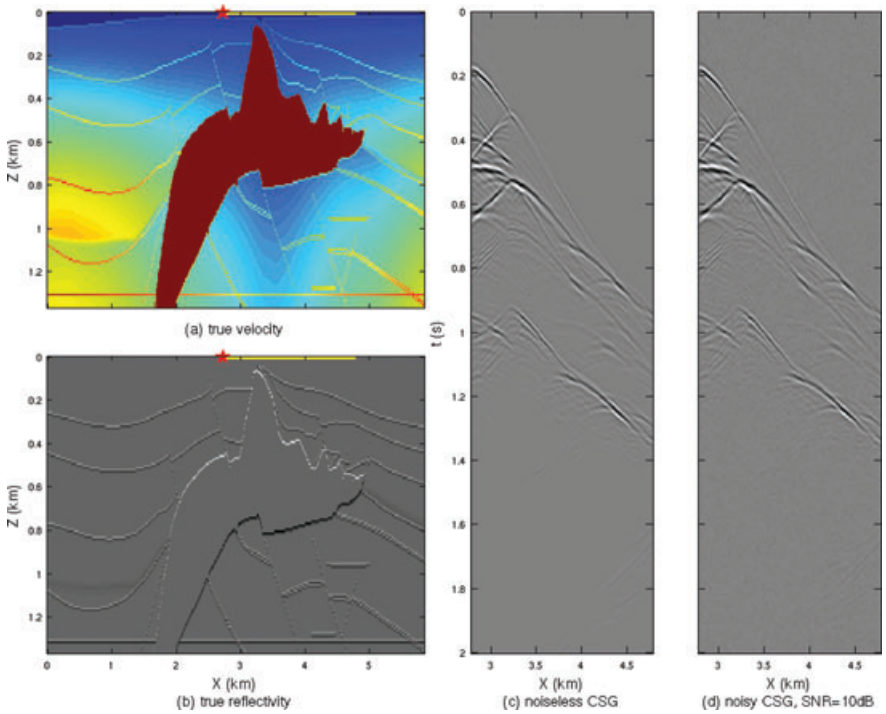

Figure 2. (a) The 2D SEG/EAGE salt model and (b) the associated reflectivity model, where the red star denotes a source at $X = 2.725$ (km) and the appending yellow line denotes the receiver aperture of this source. (c) The CSG from this source. (d) This CSG corrupted by band-limited incoherent noise such that SNR=10 dB.

computational complexities of the algorithms studied in this paper.

To demonstrate the effectiveness of LSM, its performance will be compared to that of iterative stacking (IS). In contrast to the iterative refinement of LSM, IS of encoded migration images (Schuster et al. 2011) at the k^{th} iteration produces a sum of k realizations of migration images. For IS, dynamic encoding is used, so that at each iteration the input supergather, specifically the source wavefield at surface $P(x, z = 0, \omega)$ in Fig. 8(a), is formed using a new frequency assignment.

It is of interest to analyse how the proposed method would fare compared to standard migration in terms of saving computational cost. This analysis is provided in Appendix C. In addition, the results in Appendix C allow us to compare the convergence performances of LSM and IS on the basis of the same computational cost.

RESULTS

The proposed method of Least-Squares Frequency-Division MultiSource (LSFDMS) is tested on the 2D SEG/EAGE salt model, of size[6] $n_x \times n_z = 640 \times 150$, with a grid spacing of 9.144 m. The velocity and the reflectivity model are shown in Fig. 2(a,b), respectively.

The following parameters are chosen to emulate marine acquisition geometry: shot interval = 18.288 m, receiver interval = 9.144 m, near-offset = 45.72 m, line length = 2 km. The number, M_{ga}, of supergathers dividing up all $S_{tot} = 304$ sources varies from 1,2,4, up to 8. A Ricker wavelet with a 32 Hz peak frequency is used as the source wavelet and 160 frequency channels equally divide the frequency range from 0–80 Hz, as exemplified alongside equation (27). With the true velocity and reflectivity models, a CSG for the source and receivers depicted in Fig. 2(a, b) is generated for example using split-step forward modelling and is presented in Fig. 2(c). To probe noise robustness, we contaminate the CSG's with various levels of random noise for a flat spectrum below 80 Hz, to yield SNR=10, 20, 30 dB. Figure 2(d) shows a contaminated version of Fig. 2(c). The noisy CSG's are first Wiener filtered, before being migrated. The smoothed velocity model shown in Fig. 4(a) is used as the migration velocity and is obtained by applying a 3×3 boxcar filter to the true velocity model shown in Fig. 2(a).

As the LSM iterations proceed, the trial reflectivity model is updated and surpasses the standard migration image in quality, as demonstrated in Figs. 3 and 4. For comparison,

[6] n_x is reduced from the original value of 645 to speed up the FFT.

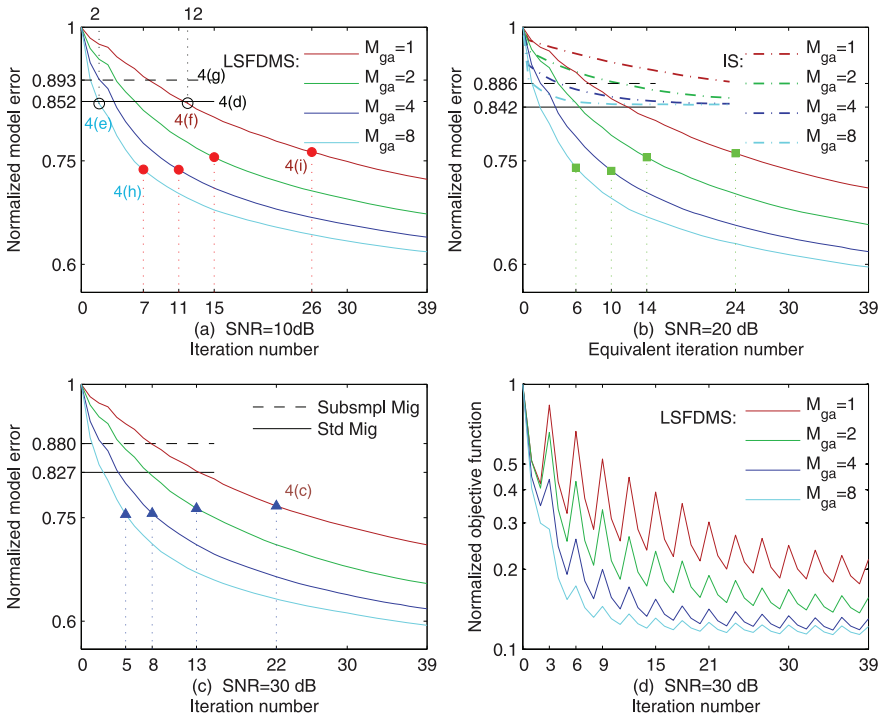

Figure 3. Normalized model error (a-c) and normalized objective function (d) for various SNR as a function of iteration number, in solid curves colour coded for various M_{ga}, when minimizing by a hybrid CG. For space efficiency, the legends in (a,c) are shared among (a-c). Regarding the black horizontal solid and dashed lines in (a-c), the dash-dot curves in (b) and the symbols ●, ■, ▲ and ○ in (a-c), see text for details. The alphanumeric short labels '4(c)' up to '4(i)' refer to in which figure and panels the corresponding migration images are shown.

migration with the subsampled CSG's ('Subsmpl Mig') is also considered, which is an alternative means for data reduction and speedup. To yield a speedup of around 8 (see Fig. 5), comparable to that of our proposed method, the subsampling ratio of 'Subsmpl Mig' is chosen as 1/8. As indicated by the black dashed horizontal lines in Fig. 3(a–c), the model error of 'Subsmpl Mig' always exceeds that of standard migration, indicated by the black solid horizontal lines. As shown in Fig. 4(g), the image produced by 'Subsmpl Mig' contains many artefacts that are disruptive because they are of similar spatial frequency and locations to those of reflectors.

Several features in Fig. 3 are worth commenting. First, understandably, larger M_{ga} and SNR lead to a smaller model error and better convergence. Second, oscillations in the objective function in panel Fig. 3(d) are the expected behaviour of a hybrid CG. The objective function is consistently reduced by CG within every $K_{CGit} = 3$ updates but will increase upon the presentation of newly encoded supergathers. This is because the previous optimization efforts are targeted at reducing a differently parametrized objective function. As the iterations proceed, however, the envelope of the oscillatory objective function still decreases, validating the robust performance of a hybrid CG.

Third, in terms of model error, the least-squares method can surpass standard migration in as few as two iterations; see for example the ○ symbol at iteration 2 on the cyan curve in Fig. 3(a). This estimate, however, is too optimistic, even though we have made sure to minimize the model error of standard migration image as $\min_\alpha \|\alpha\breve{\mathbf{m}} - \mathbf{m}\|^2$, where $\breve{\mathbf{m}}$ is the migration image and $\mathbf{m}$ is the true model. The reason is that a standard migration image tends to be smooth and the high-frequency components are suppressed. Thus the model error could be large. On the other hand, the image obtained by LSFDMS tends to be sharper, matching the true model better in terms of the L_2-norm of the model error. The downside, however, is ringy noise, as evident in the corresponding reflectivity images shown in Fig. 4(e,f). That is why it makes sense to involve human subjects in judging the quality of resulting images.

The results, as to where the LSFDMS image quality is comparable to that of standard migration, are indicated by the symbols●, ■ and ▲, of which the abscissae are also labelled, in Fig. 3(a-c), and three of such images are shown in Fig. 4(c),

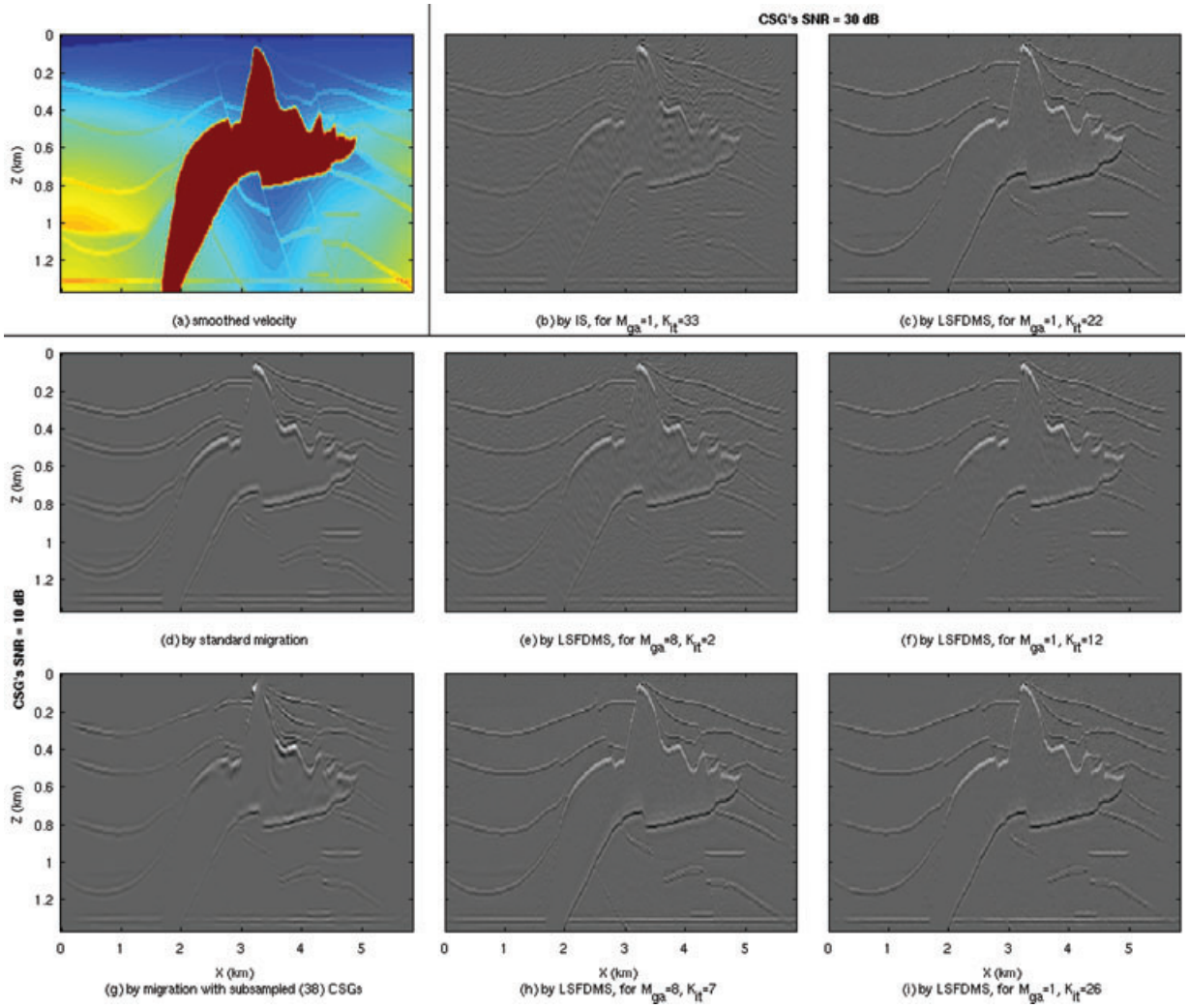

Figure 4. Reflectivity distributions obtained by various methods with a smoothed velocity model (a), in various parameter settings of M_{ga} and K_{it}, the iteration number, when applicable. 30 dB of the SNR of CSG applies to (b,c), whereas 10 dB applies to (d-i). (c–i) are respectively referred to in Fig. 3(a) and Fig. 3(c).

Fig. 4(h) and Fig. 4(i), respectively. To equate the quality of these images with that of standard migration, shown in Fig. 4(d), trade-offs are made. In Fig. 4(c), Fig. 4(h) and Fig. 4(i), there is some residual high-frequency noise, especially at shallow depths. But this noise is quite distinct from those of reflectors and thus it hardly affects the dominant features. On the other hand, the resolution of Fig. 4(c), Fig. 4(h) and Fig. 4(i) is better than that of Fig. 4(d). It is based on these two factors that we choose the break-even points in visual quality. Once the abscissae, or K_{it}'s, of these break-even points are known, from equation (C3) we calculate the relative computational cost, or, its reciprocal, termed 'gain in computational efficiency', which is plotted in Fig. 5. Here we see that, for the parameter settings and the model under study, nearly an order of magnitude of speedup can be achieved.

One may raise the concern that, due to the frequency-division scheme, even with a dozen iterations of dynamic encoding, each source can hardly have the chance to exhaust its spectrum. For example, take $M_{ga} = 2$, $K_{it} = 10$, then $S = S_{tot}/M_{ga} = 304/2 = 152$. So at any one iteration, each source only gets assigned 1/152 of the frequency channels available. With 10 iterations, in the best scenario a source can only cover a mere 10/152 of its spectrum. In light of this analysis, the apparent good performance of the frequency-division scheme seems therefore rather counter intuitive. To address this concern, we maintain that due to least-squares iterations, sources

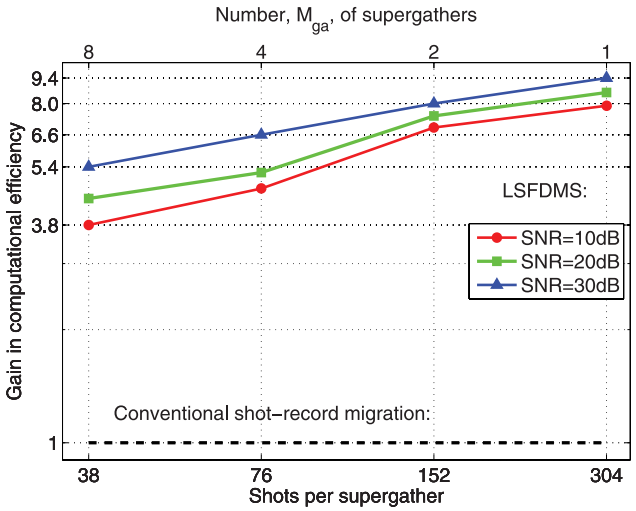

Figure 5. The gains in computational efficiency of the proposed LSFDMS as compared to the conventional shot-record split-step migration plotted on a log-log scale for various SNR's, as functions of shots per supergather, or equivalently as functions of M_{ga} (labelled at the top).

no longer act in straightforward linear superposition as they do in standard migration. Rather, they act cooperatively and with collaboration between sources the model gets effectively illuminated by a wider range of specta than provided by stacking migrations.

To test this idea, we examine the convergence performance of IS, where frequency-division encoding with multisource applies as well. Figure 3(b) includes the convergence curves (the dash-dot curves are for IS), plotted according to what is prescribed at the end of Appendix C and Fig. 4(b) shows a migration image of IS, which is obtained at the same computational cost as Fig. 4(c). Evidently, with this amount of computation, IS does not beat standard migration in terms of either model error or the quality of the migration image. The explanation for this phenomenon is precisely the concern raised earlier, aided by the realization that by random frequency assignments, rarely can there be a smooth spectrum result and fluctuations in the spectrum are likely. Non-smoothness in the spectrum corresponds to ringiness in the time domain. Therefore the migration image is always inferior to the standard migration image. Contrasting LSFDMS with IS, one can see the essential role that least-squares updates play in this frequency-division multisource method. Additional insights are reaped from a comparison study conducted in Appendix D, where we show that iterative refinement likely leads to better solutions than migration does.

To test the viability of the frequency-division multisource method in processing 3D data, we use a 3D SEG/EAGE salt model, of size $n_x \times n_y \times n_z = 672 \times 672 \times 185$ with a grid interval of 20 m. Slices of the velocity model are depicted in Fig. 6. There is one receiver at each grid point and $S_{tot} = 64 \times 64 = 4096$ sources are equally distributed on the surface. A Ricker wavelet with a 16 Hz peak frequency is used as the source wavelet; $n_\omega = 360$ frequency channels equally divide the frequency range from 0–40 Hz, as exemplified alongside equation (27). Here, fixed acquisition geometry of both the sources and receivers is assumed, as the aim of this study is to test whether the frequency-division multisource method can work on either land or marine 3D data.

Note that in this case the number of sources S_{tot} is far greater than the number of available frequency channels n_ω. If $S = S_{tot}/M_{ga} > n_\omega$, then the assignment of non-overlapping source spectra is not possible, unless only a small number of sources are turned on at a time, a practice that would discard much useful information. Here we allow overlapping source spectra. If $S \gg n_\omega$, each frequency channel is shared among S/n_ω sources. This assignment can be implemented for example by randomly drawing S/n_ω source indices in turn without replacement to be assigned to each frequency. In addition, a random polarity ± 1 is assigned to each source, in order to reduce the crosstalk among sources sharing a frequency. A comparison of this method with standard migration is given in Fig. 7, where 50 steepest descent updates of LSFDMS in one supergather yield a result comparable to standard migration. Equation (C5) says that the speedup is $2S/(4K_{it} - 1) = 2 * 4096/199 \cong 41$, if the I/O cost is ignored.

CONCLUSIONS

We emphasize that the mismatch between the limited number of live hydrophones in a marine-streamer survey and the pervasive number of live traces generated by modelling is essentially a form of crosstalk in multisource migration/inversion not seen in a fixed spread survey. To completely remove this marine crosstalk we propose a frequency-division encoding scheme, similar to the ones used in the communications industry. Our frequency-division scheme demands that, any crosstalk-prone source should only emit signals in non-overlapping frequency bands. This allows any receiver to selectively tune in to the valid source—the source that indeed has made a contribution to this receiver—and to selectively disregard potentially confounding sources; such sources are grouped with the receiver only at the time of multisource modelling.

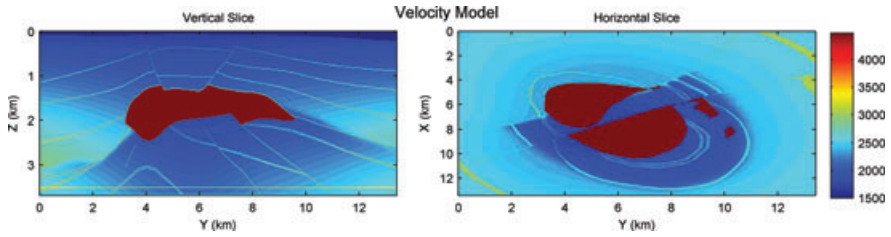

Figure 6. The 3D SEG/EAGE salt velocity model, in m/s, sliced at (left panel) x = 6.7 km and (right panel) z = 1.98 km.

We show that, our frequency-division technique in a multisource framework is similar in form to the well-studied stochastic optimization problem. In terms of computational cost, blending groups of S sources together to form supergathers would cut down the subsequent computational cost by a factor of S. Because of the weakened illumination capability however, iterations are usually required to produce an image comparable to standard migration. Fast convergence with many fewer steps than S yields an overall speed gain compared to conventional migration.

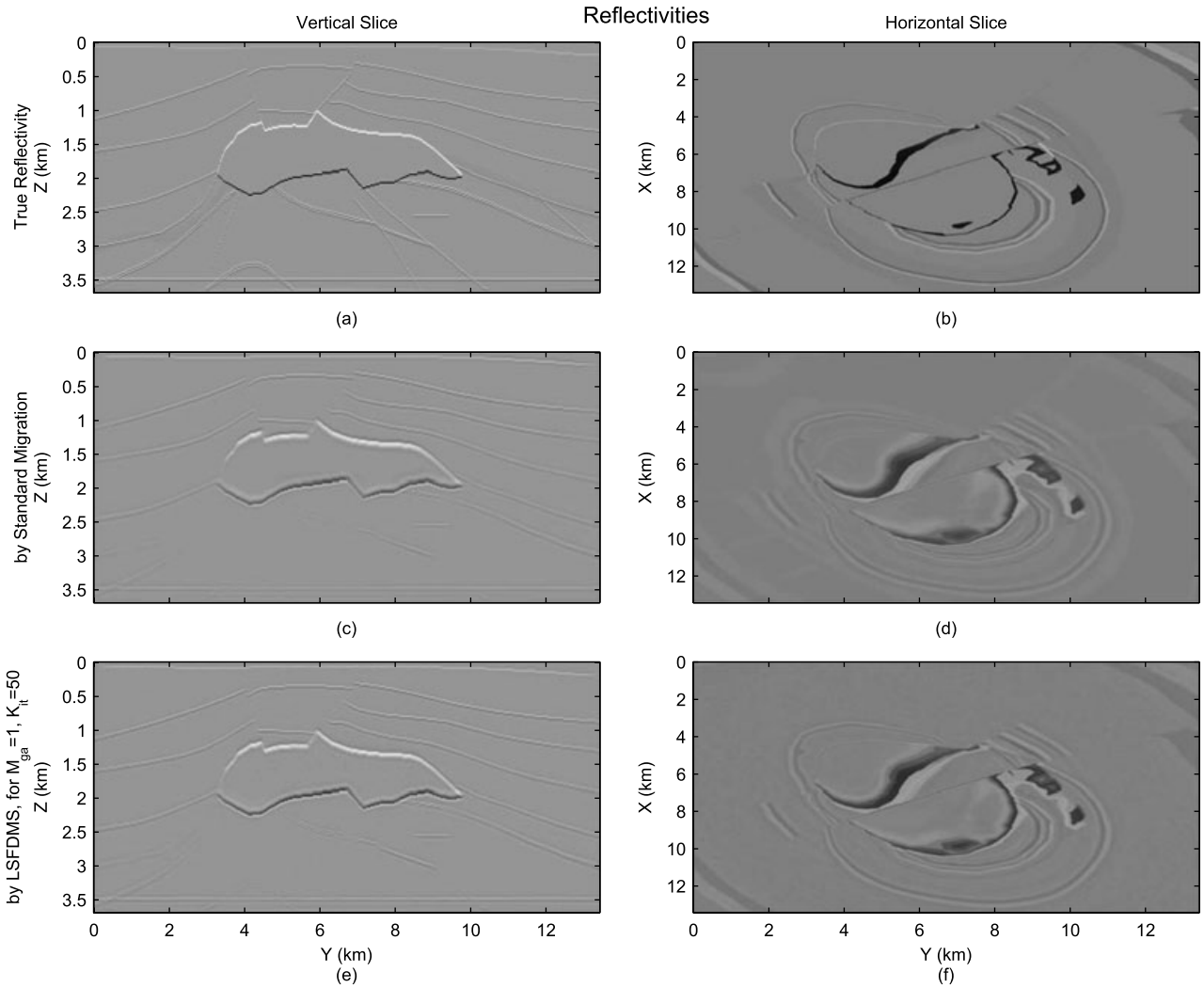

Figure 7. Reflectivity model (a,b) and others obtained by (c,d) standard shot-record prestack split-step migration and (e,f) the proposed LSFDMS for 1 supergather at the 50th iteration of steepest descent.

Numerical results with a 2D salt model and a marine survey show that crosstalk is completely removed with a multisource speedup nearly an order of magnitude faster than standard migration. In the 3D example with fixed acquisition geometry (fixed OBS geometry) a speed up of 40 was achieved compared to standard migration. In addition, better resolution was achieved. Here, because the number of available frequency channels is smaller than the number of sources, strict non-overlapping frequency assignment is impossible. To be applicable to 3D marine streamer data, we note that inline sources need to be crosstalk-free while sources along the cross-line can have crosstalk without causing the deleterious mismatch problem. This kind of crosstalk can be mitigated by existing various phase-encoding techniques. This extension is a subject of our current research.

ACKNOWLEDGEMENTS

We wish to thank the sponsors of the Center for Subsurface Imaging and Fluid Modelling (CSIM) at KAUST for their support.

REFERENCES

Amundsen L. 1991. Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. *Geophysics* **56**, 2027–2035.

Aster R., Borchers B. and Thurber C. 2005. *Parameter Estimation and Inverse Problems*. Elsevier.

Ben-Hadj-Ali H., Operto S. and Virieux J. 2009. Three-dimensional frequency-domain full waveform inversion with phase encoding. *SEG Technical Program Expanded Abstracts* **28**, 2288–2292.

Ben-Hadj-Ali H., Operto S. and Virieux J. 2011. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics* **76**, R109–R124.

Boonyasiriwat C. and Schuster G. 2010. 3d multisource full-waveform inversion using dynamic quasi-Monte Carlo phase encoding. Proceedings of the 80th Annual International Meeting, Society of Exploration Geophysicists, 1044–1049.

Dai W. and Schuster G.T. 2009. Least-squares migration of simultaneous sources data with a deblurring filter. *SEG Technical Program Expanded Abstracts* **28**, 2990–2994.

Debski W. 2010. Probabilistic inverse theory. *Advances in Geophysics* **52**, 1–102.

Duquet B., Lailly P. and Ehinger A. 2001. 3D plane wave migration of streamer data. 71st Annual International Meeting, Society of Exploration Geophysicists 1033–1036.

Duquet B., Marfurt K.J. and Dellinger J. 2000. Kirchhoff modelling, inversion for reflectivity, and subsurface illumination. *Geophysics* **65**, 1195–1209.

Gao F., Atle A. and Williamson P. 2010. Full waveform inversion using deterministic source encoding. *SEG Technical Program Expanded Abstracts* **29**.

Godwin J. and Sava P. 2010. Simultaneous source imaging by amplitude encoding. Technical Report CWP-645, Center for Wave Phenomena, Colorado School of Mines.

Jing X., Finn C.J., Dickens T.A. and Willen D.E. 2000. Encoding multiple shot gathers in prestack migration. *SEG Technical Program Expanded Abstracts* **19**, 786–789.

Krebs J.R., Anderson J.E., Hinkley D., Neelamani R., Lee S., Baumstein A. and Lacasse M.-D. 2009. Fast full-wavefield seismic inversion using encoded sources. *Geophysics* **74**, WCC177–WCC188.

Kuehl H. and Sacchi M.D. 1999. Least-squares split-step migration using the hartley transform. *SEG Technical Program Expanded Abstracts* **18**, 1548–1551.

Morton S.A. and Ober C.C. 1998. Faster shot-record migrations using phase encoding. 68th Annual International Meeting, Society of Exploration Geophysicists 1131–1134.

Nemeth T., Wu C. and Schuster G. 1999. Least-squares migration of incomplete reflection data. *Geophysics* **64**, 208–221.

Romero L.A., Ghiglia D.C., Ober C.C. and Morton S.A. 2000. Phase encoding of shot records in prestack migration. *Geophysics* **65**, 426–436.

Sacchi M. 1997. Reweighting strategies in seismic deconvolution. *Geophysical Journal International* **129**, 651–656.

Scales J., Gersztenkorn A. and Treitel S. 1988. Fast l_p solution of large, sparse, linear systems. Application to seismic travel time tomography. *Journal of Computational Physics* **75**, 314–333.

Schraudolph N.N. and Graepel T. 2002. Conjugate Directions for Stochastic Gradient Descent: ICANN. Springer Verlag, 1351–1356.

Schuster G.T., Wang X., Huang Y., Dai W. and Boonyasiriwat C. 2011. Theory of multisource crosstalk reduction by phase-encoded statics. *Geophysical Journal International* **184**, 1289C1303.

Spall J. 2003. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley and Sons.

Stoffa P.L., Fokkema J.T., de Luna Freire R.M. and Kessinger W.P. 1990. Split-step Fourier migration. *Geophysics* **55**, 410–421.

Stolt R.H. and Benson A.K. 1986. *Seismic Migration: Theory and Practice*. Pergamon Press.

Tang Y. 2009. Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian. *Geophysics* **74**, WCA95–WCA107.

Tang Y. and Biondi B. 2009. Least-squares migration/inversion of blended data. *SEG Technical Program Expanded Abstracts* **28**, 2859–2863.

Vigh D. and Starr E.W. 2008. 3d prestack plane-wave, full-waveform inversion. *Geophysics* **73**, VE135–VE144.

Virieux J. and Operto S. 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics* **74**, WCC1.

Wang J. and Sacchi M. 2007. High-resolution wave-equation amplitude variation-with-ray-parameter (AVP) imaging with sparseness constraints. *Geophysics* **72**, S11–S18.

Whitmore N.D. and Garing J.D. 1993. Interval velocity estimation using iterative prestack depth migration in the constant angle domain. *The Leading Edge* **12**, 757–762.

Zhang Y., Sun J., Notfors C., Gray S., Chernis L. and Young J. 2003. Delayed-shot 3D prestack depth migration. *SEG Technical Program Expanded Abstracts* **22**, 1027–1030.

APPENDIX A: THE EFFECT OF FREQUENCY DIVISION ON THE HESSIAN

The effect of frequency division (FD) on the Hessian matrix is now investigated. For ease of discussion, we restrict our attention to sources $s = 1, \ldots, S$ which are used to form one supergather. The Hessian can be identified as

$$\tilde{\mathbf{H}} = \sum_{j=1}^{n_\omega} \tilde{\mathbf{L}}^\dagger \tilde{\mathbf{L}} \tag{A1}$$

$$= \sum_{s=1}^{S} \sum_{j=1}^{n_\omega} N_s(j) |W(j)|^2 \underline{\mathbf{L}}_s^\dagger \underline{\mathbf{L}}_s. \tag{A2}$$

Here, equation (A2) follows from equations (8), (4), (22) and (21), and the fact that $N_s^2(j) = N_s(j)$, as $N_s(j) \in \{0, 1\}$; equation (21) ensures that all cross terms in equation (A1) when expanded by plugging in equation (8) will vanish.

In contrast, the Hessian in the standard case is

$$\mathbf{H} = \sum_{s=1}^{S} \sum_{j=1}^{n_\omega} |W(j)|^2 \underline{\mathbf{L}}_s^\dagger \underline{\mathbf{L}}_s, \tag{A3}$$

which is equation (A2) lacking the binary encoding function $N_s(j)$.

Comparing equations (A2) and (A3), we see that the encoded Hessian $\tilde{\mathbf{H}}$ consists in a subset of terms in the standard Hessian $\mathbf{H}$.

APPENDIX B: PRESTACK SPLIT-STEP MIGRATION

We describe the steps of prestack split-step migration, of which the flowcharts are illustrated in Fig. 8. This presentation closely follows Kuehl and Sacchi (1999) and is included here for convenience because analysis of computational complexity refers to it.

Consider first the forward propagation of a wavefield. The split-step operator $\mathcal{L}$ per layer can be decomposed into a succession of four linear operators $\mathcal{F}$, $\mathcal{P}$, $\mathcal{F}^{-1}$ and $\mathcal{C}$ as follows. First, the seismic wavefield $P(x, z, \omega)$ at z is transformed to the wavenumber k_x domain by the Fourier operator $\mathcal{F}$. Second, the phase-shift operator $\mathcal{P}$ is applied to the wavefield in the (k_x, ω) domain:

$$P_1(k_x, z, \omega) = P(k_x, z, \omega) e^{-i\Delta z \sqrt{(\omega u_0)^2 - k_x^2}}, \tag{B1}$$

where u_0 is the mean slowness for the current layer. Third, $P_1(k_x, z, \omega)$ is transformed to the space x domain by the inverse Fourier operator $\mathcal{F}^{-1}$. Fourth, the phase-correction operator $\mathcal{C}$ is applied in the (x, ω) domain. This accounts for the lateral slowness variation $\Delta u(x) = u(x) - u_0$:

$$P(x, z + \Delta z, \omega) = P_1(x, z, \omega) e^{-i\omega \Delta z \Delta u(x)}. \tag{B2}$$

Altogether, it is given that

$$\mathcal{L} = \mathcal{C}\mathcal{F}^{-1}\mathcal{P}\mathcal{F}, \tag{B3}$$

of which the adjoint is

$$\begin{aligned} \mathcal{L}^\dagger &= \mathcal{F}^\dagger \mathcal{P}^\dagger \mathcal{F}^{-1^\dagger} \mathcal{C}^\dagger \\ &= \mathcal{F}^{-1} \mathcal{P}^* \mathcal{F} \mathcal{C}^*, \end{aligned} \tag{B4}$$

The adjoint operator $\mathcal{L}^\dagger$ applies to the case of 'backward propagation', or downward continuation of the data, as illustrated in Fig. 8(c). This ensures that the migration operator is the adjoint of the forward modelling counterpart.

APPENDIX C: THE RELATIVE COMPUTATIONAL COST

The computational costs of the proposed method and of standard migration (and therefore the relative computational cost) will be derived as follows, assuming the observed CSG's are already in the frequency domain. Since ultimately it is the ratio between the computational costs of different methods that is of interest, we restrict our attention to one frequency and to the S sources encompassed by one supergather.

First, spatial FFT's, repeatedly invoked in Fig. 8(a–c), dominate the computational cost of split-step migration. Elementwise product and dot product between two vectors of length n_x, the grid size along x, incur a cost C_p that is a small multiple of n_x and the number of times such computation is invoked is comparable to that of FFT_x. On the other hand, the cost of each $\text{FFT}_x \cong 4n_x \log_2(n_x)$. For typical seismic reflectivity models, $n_x > 512$ and thus this cost $\gtrsim 36 n_x n_z n_\omega$, far exceeding C_p. Since every flowchart in Fig. 8 contains an equal number of FFT_x's, from now on we take the computational cost for each of these flowcharts as of unit 1.

Next, the computational costs for forward modelling and for migration, in cases of whether the source field is available, are given by the following *items*:

1. Forward modelling

a. The source field, $P(x, z, \omega)$ as depicted in Fig. 8(a), is not available. Therefore $P(x, z, \omega)$ needs to be computed before the reflected field $R(x, z, \omega)$, as depicted in Fig. 8(b), can be obtained. The computational cost is thus 2.

b. The source field is available. Then the only task is to compute $R(x, z, \omega)$, with a computational cost of 1.

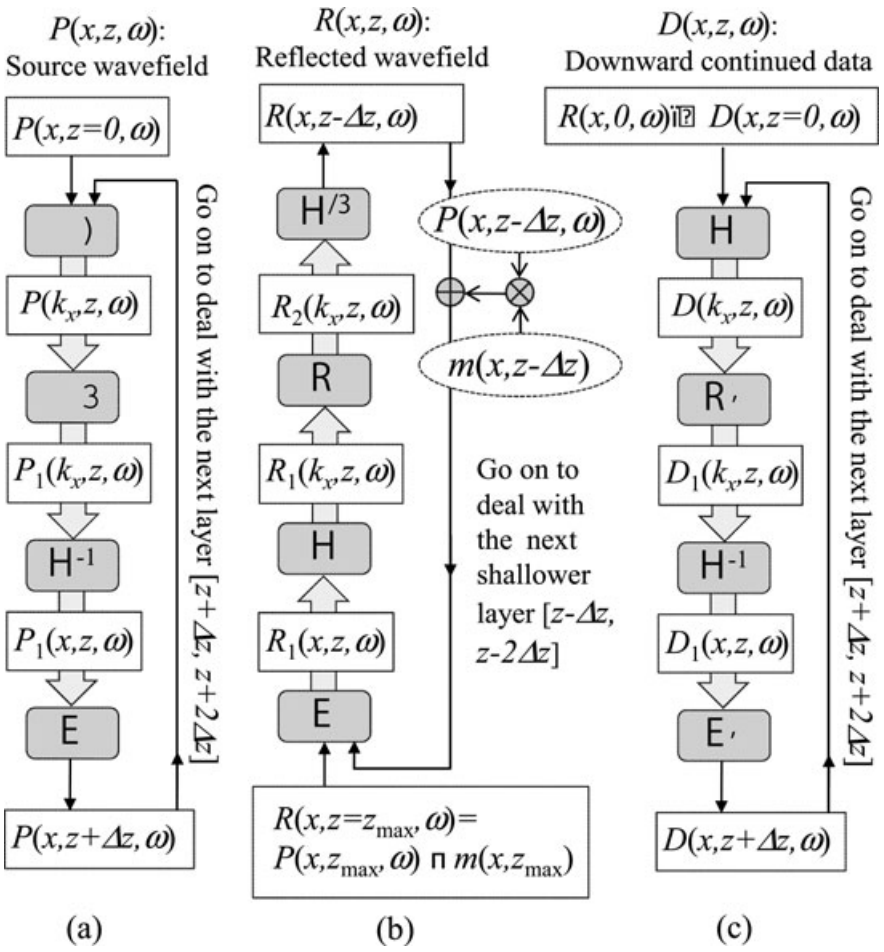

Figure 8. Flowcharts for prestack split-step modelling and migration. $\mathcal{F}$, $\mathcal{F}^{-1}$, $\mathcal{P}$ and $\mathcal{C}$ denote the Fourier transform, the inverse Fourier transform, the phase-shift operator and the phase-correction operator, respectively (see text for details). (a) The source wavefield $P(x, z, \omega)$ is propagated from the surface of the earth to depth z in steps Δz. (b) At each depth z of the earth, a reflected wave is generated by $m(x, z)P(x, z, \omega)$, where $m(x, z)$ is a reflectivity model. The wave is then propagated upward to the surface $z = 0$. The total reflected wavefield $R(x, z, \omega)$ consists in the superposition of the reflected and propagated waves originating from below. The total reflected wavefield collected at the surface are the data, i.e., $R(x, z = 0, \omega) \equiv D(x, z = 0, \omega)$. (c) The data are then downward continued from the surface back to depth z in steps Δz. Finally, the migration image $I(x, z)$ (not shown) is constructed by applying the imaging condition: $I(x, z) = \sum_\omega P^*(x, z, \omega)D(x, z, \omega)$, or $I(x, z) = 2\sum_{\omega>0}\mathrm{Re}\{P^*(x, z, \omega)D(x, z, \omega)\}$, assuming the DC component is 0.

2. Migration

a. The source field is not available. Then both the source field and the downward continued data field $D(x, z, \omega)$, as depicted in Fig. 8(c), need to be computed. The computational cost is 2. Note: this applies to standard migration.

b. The source field is available. Then only the downward continued data field needs to be computed, before forming the final migration image. This computational cost is 1.

With these results in mind, we study the computational cost incurred in a CG algorithm, as listed in Algorithm 1. For standard migration, there are S sources covered by the supergather in question and therefore the computational cost as remarked at the end of item 2(a) needs a factor S. Let κ_0 be this cost, expressed as

$$\kappa_0 = 2S, \tag{C1}$$

and let κ_{LS} be the cost for LSM, given in the comment besides line 25 as

$$\kappa_{LS} = K_{it}\frac{3 + 2K_{CGit}}{K_{CGit}} - 1. \tag{C2}$$

Therefore the relative computational cost is given by

$$\rho = \frac{\kappa_{LS}}{\kappa_0} = \frac{K_{it}}{S}\frac{3 + 2K_{CGit}}{2K_{CGit}} - \frac{1}{2S} \tag{C3}$$

Algorithm 1. Conjugate Gradient algorithm, for solving $\mathbf{Lm} = \mathbf{d}$.

1: **procedure** CONJGRAD($\mathbf{m}$, $\mathbf{L}$, $\mathbf{d}$, K_{CGit})
2: **if** $\mathbf{m} = 0$ **then**
3: $\mathbf{g} \leftarrow \mathbf{L}\dagger\mathbf{d}$ ▷ item 2(a) ⇒ cost=2.
4: **else**
5: $\mathbf{g} \leftarrow -\mathbf{L}\dagger(\mathbf{Lm} - \mathbf{d})$ ▷ item 1(a) and 2(b) ⇒ cost=2+1. In forward modelling, the source field is not available, ∴ item 1(a); in migration, the source field has just become available, ∴ item 2(b).
6: **end if**
7: $\mathbf{p} \leftarrow \mathbf{g}$
8: $g2_{old} \leftarrow ||\mathbf{g}||^2$
9: **if** $K_{CGit} = 1$ ▷ Steepest Descent
10: $\alpha \leftarrow \frac{g2_{old}}{||\mathbf{Lp}||^2}$ ▷ Forward modelling. Source fields have been computed in line 3 or 5. ∴ Item 1(b) ⇒ cost=1.
11: $\mathbf{m} \leftarrow \mathbf{m} + \alpha\mathbf{p}$
12: **return m** For steepest descent, if this procedure is called K_{it} times, the total cost is $3 + 4(K_{it} - 1) = 4K_{it} - 1$.
13: **else** ▷ Conjugate Gradient
14: **for** $k \leftarrow 1, K_{CGit}$
15: $\mathbf{q} \leftarrow \mathbf{L}\dagger(\mathbf{Lp})$ ▷ Forward modelling and migration. Since among CG updates, the sources are fixed rather than encoded anew, source fields are thus fixed and have been computed in line 3 or 5. As the trial reflectivity model $\mathbf{m}$ is being updated, item 1(b) applies. For migration, item 2(b) applies. So the total cost for this line is 2.
16: $\alpha \leftarrow \frac{g2_{old}}{\mathbf{p}^\dagger\mathbf{q}}$
17: $\mathbf{m} \leftarrow \mathbf{m} + \alpha\mathbf{p}$
18: $\mathbf{g} \leftarrow \mathbf{g} - \alpha\mathbf{q}$
19: $g2_{new} \leftarrow ||\mathbf{g}||^2$
20: $\mathbf{p} \leftarrow \mathbf{g} + \frac{g2_{new}}{g2_{old}}\mathbf{p}$
21: $g2_{old} \leftarrow g2_{new}$
22: **endfor** ▷ Taken together, for this loop, the total cost is $2K_{CGit}$.
23: **return m** ▷ As this procedure is called K_{CG} times, starting from model 0, the total computational cost is $\kappa_{LS} = 2 + 2K_{CGit} + (K_{CG} - 1)(3 + 2K_{CGit}) = K_{CG}(3 + 2K_{CGit}) - 1$.
24: **end if**
25: **end procedure** ▷ In an alternative formulation, take $K_{it} = K_{CG}K_{CGit}$ and express κ_{LS} in terms of K_{it} as $\kappa_{LS} = K_{it}\frac{3+2K_{CGit}}{K_{CGit}} - 1$.

$$= \frac{3K_{it} - 1}{2S}, \text{ when } K_{CGit} = 3. \tag{C4}$$

In the case of steepest descent, the relative computational cost is given by

$$\bar{\rho} = \frac{4K_{it} - 1}{2S}, \tag{C5}$$

which follows from the comment regarding line 1.12 and equation (C1).

In the case of IS with encoded supergathers, the computational cost per iteration is equal to that of standard migration, except without the S factor. For J_{it} stackings, the cost is thus

$$\kappa_{IS} = 2J_{it}. \tag{C6}$$

Equating equations (C2) and (C6) leads to

$$J_{it} = K_{it}\frac{3 + 2K_{CGit}}{2K_{CGit}} - \frac{1}{2} \tag{C7}$$

$$= \frac{3K_{it} - 1}{2}, \quad \text{when } K_{CGit} = 3. \tag{C8}$$

Equation (C7) relates J_{it}, the number of iterations of IS, to K_{it}, the number of iterations of LSM, on the condition that the computational costs incurred by these two methods are equal. This allows us to compare two criterion functions for LSM and IS, respectively, on the basis of the same computational cost. Let $f_{LSM}(K_{it})$ and $f_{IS}(J_{it})$ be criterion functions of iteration step K_{it} and J_{it}, for LSM and IS, respectively. Using equation (C7), we write $f_{IS}(J_{it}) = f_{IS}(J_{it}(K_{it}))$. Plotted on the abscissa of K_{it}, the two curves of $f_{LSM}(K_{it})$ and $f_{IS}(J_{it}(K_{it}))$ are thus put on equal footing of computational cost. Examples of such plots are provided in Fig. 3(b).

APPENDIX D: MIGRATION VERSUS SUCCESSIVE STEEPEST DESCENT

Given a set of sampled Hessians (or equivalently sampled modelling operators) and the associated data generated according to an underlining model $\mathbf{m}$, we investigate two strategies, with the knowledge of these sampled Hessians and observed data, to get an estimate of $\mathbf{m}$: 1) migration and 2) successive steepest descent (SSD). The question we intend to address is, which strategy is better—in terms of producing a more accurate estimate?

Let $\mathbf{L}_1, \ldots, \mathbf{L}_n$ and $\mathbf{d}_1, \ldots, \mathbf{d}_n$ respectively be the n copies of sampled modelling operators and the data generated according to

$$\mathbf{d}_i = \mathbf{L}_i \mathbf{m}, \quad \forall i = 1, \ldots, n. \tag{D1}$$

The migration strategy forms the estimate as

$$\mathbf{m}_{mig} = \sum_{i=1}^{n} \mathbf{L}_i{}^{\dagger} \mathbf{d}_i. \tag{D2}$$

On the other hand, the SSD strategy consists in the following steps:

1 Start from $\mathbf{m}_{SSD}^{(0)} = \mathbf{0}$.

2 At the k^{th} step, presented with $\mathbf{L}_k$ and $\mathbf{d}_k$, update the trial model once, according to the steepest descent formula, prescribed as :

$$\mathbf{m}_{SSD}^{(k+1)} = \mathbf{m}_{SSD}^{(k)} + \alpha \mathbf{g}, \tag{D3}$$

where

$$\mathbf{g} = \mathbf{L}_k{}^{\dagger}(\mathbf{d}_k - \mathbf{m}_{SSD}^{(k)}), \tag{D4}$$

$$\alpha = \frac{||\mathbf{g}||^2}{||\mathbf{L}_k \mathbf{g}||^2}. \tag{D5}$$

3 Finish at $k = n$.

To fairly evaluate the model error, we introduce

$$\theta^{\epsilon}(\mathbf{x}) \stackrel{\text{def}}{=} \measuredangle(\mathbf{x}, \mathbf{m}) = \arccos\left(\frac{\mathbf{x}^T \cdot \mathbf{m}}{||\mathbf{x}||\ ||\mathbf{m}||}\right), \tag{D6}$$

a criterion that ignores the magnitudes of the vectors in question.

Unable to establish a mathematical bound on $\theta^{\epsilon}(\mathbf{m}_{mig})$ in comparison to $\theta^{\epsilon}(\mathbf{m}_{SSD})$, we resort to the Monte Carlo method instead. In this study, $\mathbf{d} \in \mathbb{R}^{10}$, $\mathbf{m} \in \mathbb{R}^{25}$; the sizes are deliberately chosen to make each individual inverse problem, i.e., to invert equation (D1), underdetermined. In addition, the performance of the two strategies in the presence of noise is probed. Specifically, a line of code

$$\mathbf{d}_i \leftarrow \mathbf{d}_i + \mathbf{r} \tag{D7}$$

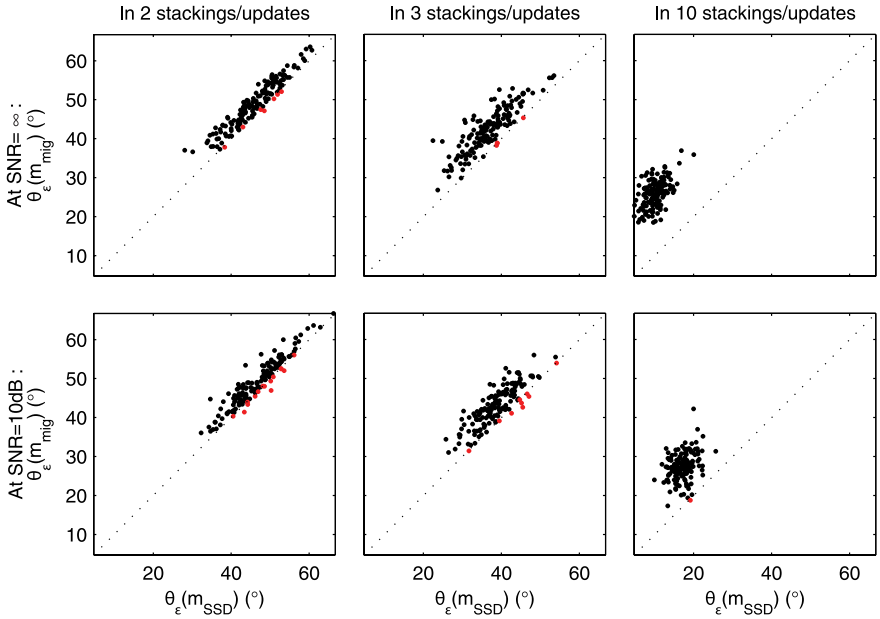

Figure 9. Scatter plot of the model errors in terms of θ^{ϵ}, for migration and for SSD, when (upper row) data are not contaminated by noise and (lower row) SNR=10dB, with the set size n varying from (left column) 2, to (middle column) 3 and to (right column) 10. In all the panels, data points corresponding to $\theta^{\epsilon}(\mathbf{m}_{mig}) >= \theta^{\epsilon}(\mathbf{m}_{SSD})$ are plotted in black, while points corresponding to $\theta^{\epsilon}(\mathbf{m}_{mig}) < \theta^{\epsilon}(\mathbf{m}_{SSD})$ are plotted in red.

follows immediately after equation (D1), and all succeeding operations are based on the noisy $\mathbf{d}_i$'s. Here, $\mathbf{r}$ is a vector of white Gaussian noise, with its power adjusted to meet a given choice of SNR.

The results are summarized in Fig. 9. Here, $\mathbf{L}_i$'s and $\mathbf{m}$ are generated using Gaussian distribution. We have varied the types of distribution, from Gaussian to a sparse distribution such as binomial, and qualitatively the same trends are observed: SSD produces a smaller model error than migration does, except for a few rare exceptions. As the size of the random set grows (i.e., more stackings in migration and more updates in SSD), the advantage of SSD becomes even more apparent. As plotted in the upper right panel, $\theta^{\epsilon}(\mathbf{m}_{mig})/\theta^{\epsilon}(\mathbf{m}_{SSD}) \approx 24/6$. When $\mathbf{d}_i$'s are corrupted by noise, however, the performance of SSD deteriorates more than that of migration does. But at this level of noise, in most cases, SSD still outperforms migration.

These observations are intuitively understandable. Migration can be thought of as one step of steepest descent starting from $\mathbf{0}$. Over a set of n samples, migration amounts to averaging n attempts of steepest descent, each starting from $\mathbf{0}$, whereas in SSD, the trial model keeps improving. So it is very probable that the latter outperforms the former. It could happen that the average of these 'first attempts' comes very close to the true model. This explains why exceptions exist. In the presence of white noise, averaging with equal weight over random instances, as migration does, is the most effective means to reduce noise. In SSD, however, the earlier updates influence the iteration trajectory more than the later updates do. In effect, the end result senses a weighted average of the noise contained in each update, with a larger weight assigned to early samples and a smaller weight to later samples, resulting in less noise reduction than what migration is capable of.