Geophysical Prospecting

EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Geophysical Prospecting doi: 10.1111/j.1365-2478.2012.01064.x

Seismic wave extrapolation using lowrank symbol approximation

Sergey Fomel[1]*, Lexing Ying[2] and Xiaolei Song[1]

[1]*Bureau of Economic Geology, John A. and Katherine G. Jackson School of Geosciences, The University of Texas at Austin, University Station, Box X Austin, TX 78713-8924, USA,* [2]*Department of Mathematics, The University of Texas at Austin, 1 University Station, Austin, TX 78712, USA*

Received January 2011, revision accepted January 2012

ABSTRACT

We consider the problem of constructing a wave extrapolation operator in a variable and possibly anisotropic medium. Our construction involves Fourier transforms in space combined with the help of a lowrank approximation of the space-wavenumber wave-propagator matrix. A lowrank approximation implies selecting a small set of representative spatial locations and a small set of representative wavenumbers. We present a mathematical derivation of this method, a description of the lowrank approximation algorithm and numerical examples that confirm the validity of the proposed approach. Wave extrapolation using lowrank approximation can be applied to seismic imaging by reverse-time migration in 3D heterogeneous isotropic or anisotropic media.

Keywords: Decomposition, Lowrank.

INTRODUCTION

Wave extrapolation in time plays an important role in seismic imaging (reverse-time migration), modelling and full waveform inversion. Conventionally, extrapolation in time is performed by finite-difference methods (Etgen 1986). Spectral methods (Tal-Ezer, Kosloff and Koren 1987; Reshef *et al.* 1988) have started to gain attention recently and to become feasible in large-scale 3D applications thanks to the increase in computing power. The attraction of spectral methods is in their superb accuracy and, in particular, in their ability to suppress dispersion artefacts (Chu and Stoffa 2008; Etgen and Brandsberg-Dahl 2009).

Theoretically, the problem of wave extrapolation in time can be reduced to analysing numerical approximations to the mixed-domain space-wavenumber operator (Wards, Margrave and Lamoureux, 2008). In this paper, we propose a systematic approach to designing wave extrapolation operators by approximating the space-wavenumber matrix symbol with a lowrank decomposition. A lowrank approximation implies selecting a small set of representative spatial locations and a small set of representative wavenumbers. The optimized separable approximation or OSA (Song 2001) was previously employed for wave extrapolation (Zhang and Zhang 2009; Du, Fletcher and Fowler 2010) and can be considered as another form of lowrank decomposition. However, the decomposition algorithm in OSA is significantly more expensive, especially for anisotropic wave propagation, because it involves eigenfunctions rather than rows and columns of the original extrapolation matrix. Our algorithm can also be regarded as an extension of the wavefield interpolation algorithm of Etgen and Brandsberg-Dahl (2009), with optimally selected reference velocities and weights. Another related method is the Fourier finite-difference (FFD) method proposed recently by Song and Fomel (2011). FFD may have an advantage in efficiency, because it uses only one pair of multidimensional forward and inverse FFTs (fast Fourier transforms) per time step. However, it does not offer flexible controls on the approximation accuracy.

Our approach to wave extrapolation is general and can apply to different types of waves, including both acoustic and elastic seismic waves, as well as velocity continuation (Fomel 2003b), offset continuation (Fomel 2003a), prestack

*Email: sergey.fomel@beg.utexas.edu, lexing@math.utexas.edu

exploding reflector extrapolation (Alkhalifah and Fomel 2010), etc.

The paper is organized as follows. We first present the theory behind the proposed algorithm, then describe the algorithm and test its accuracy on a number of synthetic benchmark examples of increasing complexity.

WAVE EXTRAPOLATION

Let $P(\mathbf{x}, t)$ be the seismic wavefield at location $\mathbf{x}$ and time t. The wavefield at the next time step $t + \Delta t$ can be approximated by the following mixed-domain operator (Wards *et al.* 2008)

$$P(\mathbf{x}, t + \Delta t) = \int \widehat{P}(\mathbf{k}, t)\, e^{i\,\phi(\mathbf{x},\mathbf{k},\Delta t)}\, d\mathbf{k}\ , \tag{1}$$

and $\widehat{P}(\mathbf{k}, t)$ is the spatial Fourier transform of $P(\mathbf{x}, t)$

$$\widehat{P}(\mathbf{k}, t) = \frac{1}{(2\,\pi)^3} \int P(\mathbf{x}, t) e^{-i\,\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}\ , \tag{2}$$

and $\mathbf{k}$ is the spatial wavenumber. To define the phase function $\phi(\mathbf{x}, \mathbf{k}, t)$, which appears in equation (1), one can substitute approximation (1) into the wave equation and extract the geometrical (high-frequency) asymptotic of it. In case of seismic wave propagation, this leads to the eikonal-like equation

$$\frac{\partial \phi}{\partial t} = \pm V(\mathbf{x}, \mathbf{k})\ |\nabla \phi|\ , \tag{3}$$

where $V(\mathbf{x}, \mathbf{k})$ is the phase velocity, and the choice of the sign corresponds, in the case of a point source, to expanding or contracting waves. In the isotropic case, V does not depend on $\mathbf{k}$. The initial condition for equation (3) is

$$\phi(\mathbf{x}, \mathbf{k}, 0) = \mathbf{k} \cdot \mathbf{x}\ , \tag{4}$$

which turns equation (1) into the simple inverse Fourier transform operation.

Assuming small steps Δt in equation (1), one can build successive approximations for the phase function ϕ by expanding it into a Taylor series. In particular, let us represent the phase function as

$$\phi(\mathbf{x}, \mathbf{k}, t) \approx \mathbf{k} \cdot \mathbf{x} + \phi_1(\mathbf{x}, \mathbf{k})\, t + \phi_2(\mathbf{x}, \mathbf{k})\, \frac{t^2}{2} + \cdots \tag{5}$$

Correspondingly,

$$|\nabla \phi| \approx |\mathbf{k}| + \frac{\nabla \phi_1 \cdot \mathbf{k}}{|\mathbf{k}|}\, t + O(t^2)\ . \tag{6}$$

Substituting expansions (5) and (6) into equation (3) and separating the terms with different powers of t, we find that

$$\phi_1(\mathbf{x}, \mathbf{k}) = V(\mathbf{x}, \mathbf{k})\, |\mathbf{k}|\ , \tag{7}$$

$$\phi_2(\mathbf{x}, \mathbf{k}) = V(\mathbf{x}, \mathbf{k})\, \nabla V \cdot \mathbf{k}\ . \tag{8}$$

When either the velocity gradient ∇V or the time step Δt are small, Taylor expansion (5) can be reduced to only two terms, which in turn reduces equation (1) to the familiar expression (Etgen and Brandsberg-Dahl 2009)

$$P(\mathbf{x}, t + \Delta t) \approx \int \widehat{P}(\mathbf{k}, t)\, e^{i\,[\mathbf{k}\cdot\mathbf{x} + V(\mathbf{x},\mathbf{k})\,|\mathbf{k}|\,\Delta t]}\, d\mathbf{k}\ , \tag{9}$$

or

$$P(\mathbf{x}, t + \Delta t) + P(\mathbf{x}, t - \Delta t) \approx 2 \int \widehat{P}(\mathbf{k}, t)\, e^{i\,\mathbf{k}\cdot\mathbf{x}} \cos\left[V(\mathbf{x}, \mathbf{k})\, |\mathbf{k}|\, \Delta t\right]\, d\mathbf{k}\ . \tag{10}$$

In rough velocity models, where the gradient ∇V does not exist, one can attempt to solve eikonal equation (3) numerically or to apply approximations other than Taylor expansion (5). In the examples of this paper, we used only the ϕ_1 term.

Note that the approximations that we use, starting from equation (1), are focused primarily on the phase of wave propagation. As such, they are appropriate for seismic migration but not necessarily for accurate seismic modelling, which may require taking account of amplitude effects caused by variable density and other elastic phenomena.

The computational cost for a straightforward application of equation (1) is $O(N_x^2)$, where N_x is the total size of the three-dimensional $\mathbf{x}$ grid. Even for modest-size problems, this cost is prohibitively expensive. In the next section, we describe an algorithm that reduces the cost to $O(M\, N_x \log N_x)$, where M is a small number.

LOWRANK APPROXIMATION

The key idea of the lowrank decomposition is decomposing the wave extrapolation matrix

$$W(\mathbf{x}, \mathbf{k}) = e^{i\,[\phi(\mathbf{x},\mathbf{k},\Delta t) - \mathbf{k}\cdot\mathbf{x}]} \tag{11}$$

for a fixed Δt into a separated representation

$$W(\mathbf{x}, \mathbf{k}) \approx \sum_{m=1}^{M} \sum_{n=1}^{N} W(\mathbf{x}, \mathbf{k}_m) a_{mm}\, W(\mathbf{x}_n, \mathbf{k}). \tag{12}$$

Representation (12) speeds up the computation of $P(\mathbf{x}, t + \Delta t)$ since

$$\begin{aligned} P(\mathbf{x}, t + \Delta t) &= \int e^{i\mathbf{x}\mathbf{k}}\, W(\mathbf{x}, \mathbf{k}) \widehat{P}(\mathbf{k}, t) d\mathbf{k} \\ &\approx \sum_{m=1}^{M} W(\mathbf{x}, \mathbf{k}_m) \left(\sum_{n=1}^{N} a_{mm} \left(\int e^{i\mathbf{x}\mathbf{k}}\, W(\mathbf{x}_n, \mathbf{k}) \widehat{P}(\mathbf{k}, t) d\mathbf{k} \right) \right)\ . \end{aligned} \tag{13}$$

The evaluation of the last formula is effectively equivalent to applying N inverse fast Fourier transforms. Physically, a separable lowrank approximation amounts to selecting a set of N representative spatial locations and M representative wavenumbers.

In order to discuss the construction of approximation (12), let us view it as a matrix decomposition problem

$$\mathbf{W} \approx \mathbf{W}_1 \, \mathbf{A} \, \mathbf{W}_2 \tag{14}$$

where $\mathbf{W}$ is the $N_x \times N_x$ matrix with entries $W(\mathbf{x}, \mathbf{k})$, $\mathbf{W}_1$ is the submatrix of $\mathbf{W}$ that consists in the columns associated with $\{\mathbf{k}_m\}$, $\mathbf{W}_2$ is the submatrix that consists in the rows associated with $\{\mathbf{x}_n\}$ and $\mathbf{A} = \{a_{mm}\}$. In practice, we find that matrix $\mathbf{W}$ has a low-rank separated representation provided that Δt is sufficiently small, which, in the case of smooth models, can be partially explained by the separation of terms in Taylor series (5). The construction of the separated representation in equation (14) follows the method of Engquist and Ying (2007, 2009) and is detailed in the appendix. The main observation is that the columns of $\mathbf{W}_1$ and the rows of $\mathbf{W}_2$ should span the column space and row space of $\mathbf{W}$, respectively, as well as possible. Let ε be a prescribed accuracy of this separated representation and r_ε be the numerical rank of $\mathbf{W}$. The algorithm for computing equation (14) takes the following steps:

(1.) Pick a *uniformly random* set S of $\beta \cdot r_\varepsilon$ columns of $\mathbf{W}$ where β is chosen to be 3 or 4 in practice. Perform the pivoted QR factorization of $(\mathbf{W}(:, S))^*$ (Golub and Van Loan (1996). The first r_ε pivoted columns correspond to r_ε rows of the matrix $\mathbf{W}(:, S)$. Define $\mathbf{W}_1$ to be the submatrix of $\mathbf{W}$ that consists in these rows and set $\mathbf{x}_1, \ldots, \mathbf{x}_N$ with $n = r_\varepsilon$ to be the corresponding x values of these rows.

(2.) Pick a *uniformly random* set T of $\beta \cdot r_\varepsilon$ rows of $\mathbf{W}$ and perform the pivoted QR factorization of $\mathbf{W}(T, :)$. Define $\mathbf{W}_2$ to be the submatrix of $\mathbf{W}$ that consists in these columns and set $\mathbf{k}_1, \ldots, \mathbf{k}_M$ with $m = r_\varepsilon$ to be the corresponding k values of these columns.

(3.) Set the middle matrix $\mathbf{A} = \mathbf{W}^\dagger(\mathbf{x}_n, \mathbf{k}_m)_{1 \le n \le N, 1 \le m \le M}$ where $\dagger$ stands for the pseudoinverse.

(4.) Combine the result of the previous three steps to obtain the required separated representation $\mathbf{W} \approx \mathbf{W}_1 \, \mathbf{A} \, \mathbf{W}_2$.

The algorithm does not require, at any step, access to the full matrix $\mathbf{W}$, only to its selected rows and columns. Once the decomposition is complete, it can be used at every time step during the wave extrapolation process. In multiple-core implementations, the matrix operations in equation (12) are easy to parallelize. The algorithm details are outlined in the appendix.

The cost of the algorithm is $O(M N_x \log N_x)$ operations per time step, where $N_x \log N_x$ refers to the cost of the Fourier transform. In comparison, the cost of finite-difference wave extrapolation is $O(L N_x)$, where L is the size of the finite-difference stencil. Song *et al.* (2011) presented an application of the proposed lowrank approximation algorithm for devising accurate finite-different schemes. There is a natural trade-off in the selection of M: larger values lead to a more accurate wave representation but require a longer computational time. In the examples of the next section, we select these parameters based on an estimate of the approximation accuracy and generally aiming for the relative accuracy of 10^{-4}. The resulting M is typically smaller than the number of Fourier transforms required for pseudo-spectral algorithms such as pseudo-spectral implementations of the rapid expansion method (Pestana and Stoffa 2011).

EXAMPLES

We start with a simple 1D example. The 1D velocity model contains a linear increase in velocity, from 1 km/s to 2.275 km/s. The extrapolation matrix, $2\,(\cos[V(x)\,|k|\,\Delta t] - 1)$, or pseudo-Laplacian in the terminology of Etgen and Brandsberg-Dahl (2009), for the time step $\Delta t = 0.001$ s is plotted in Fig. 1(a). Its lowrank approximation is shown in Fig. 1(b) and corresponds to $N = M = 2$. The x locations selected by the algorithm correspond to velocities of 1.59 and 2.275 km/s. The wavenumbers selected by the algorithm correspond to the Nyquist frequency and 0.7 of the Nyquist frequency. The approximation error is shown in Fig. 1(c). The relative error does not exceed 0.34%. Such a small approximation error results in accurate wave extrapolation, which is illustrated in Fig. 2. The extrapolated wavefield shows a negligible error in wave amplitudes, as demonstrated in Fig. 2(c).

Our next example (Figs 3 and 4) corresponds to wave extrapolation in a 2D smoothly variable isotropic velocity field. As shown by Song and Fomel (2011), the classic finite-difference method (second-order in time, fourth-order in space) tends to exhibit dispersion artefacts in this example with the chosen model size and extrapolation step, while spectral methods exhibit high accuracy. As yet another spectral method, the lowrank approximation is highly accurate. The wavefield snapshot, shown in Fig. 3(b) and Fig. 4(b), is free from dispersion artefacts and demonstrates high accuracy. The approximation rank decomposition in this case is $N = M = 2$, with the expected error of less than 10^{-4}. In our implementation, the CPU time for finding the lowrank ap-

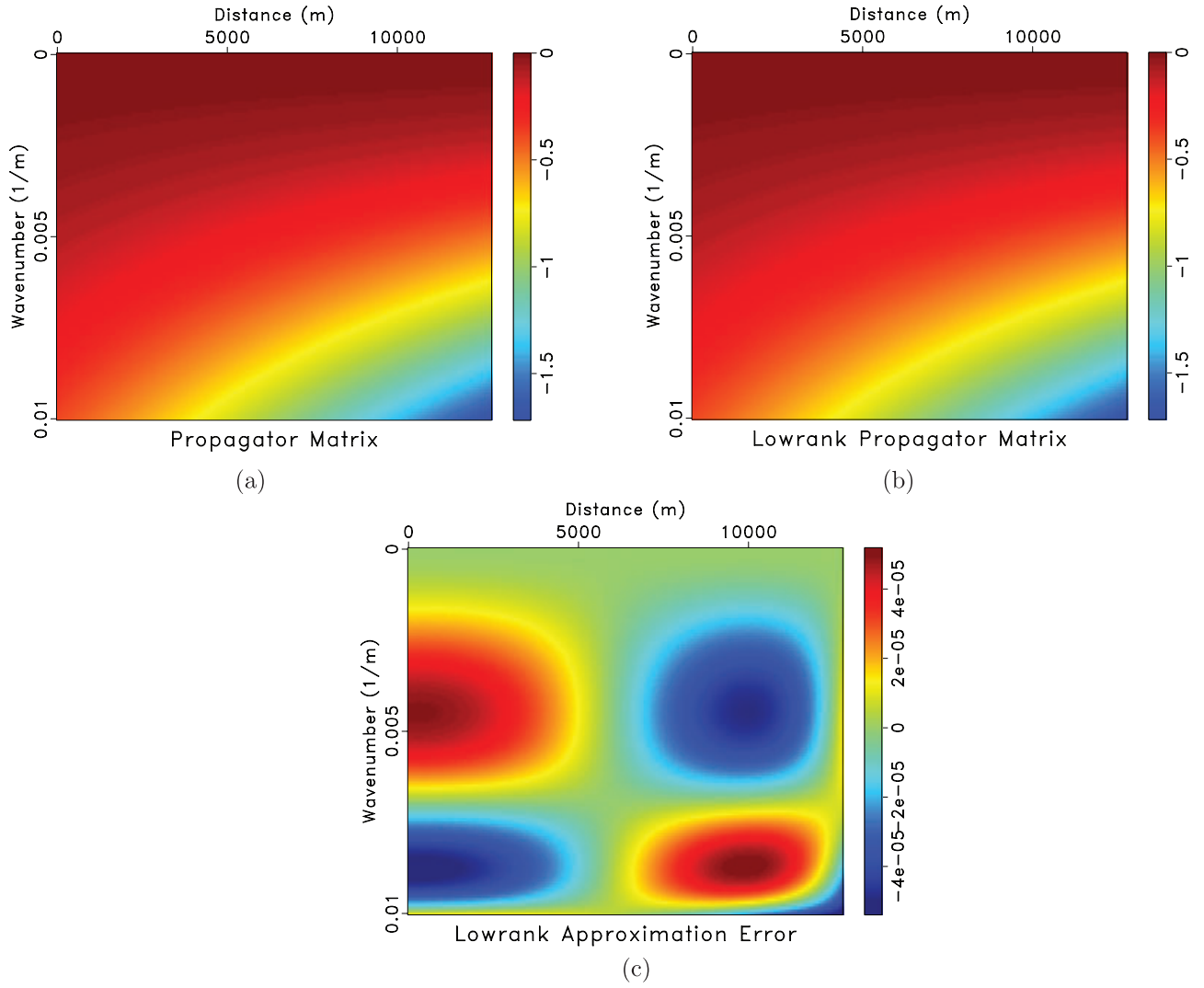

Figure 1 Wave extrapolation matrix for 1D wave propagation with linearly increasing velocity (a), its lowrank approximation (b) and approximation error (c).

proximation was 2.45 s and the single-processor CPU time for extrapolation for 2500 time steps was 101.88 s or 2.2 times slower than the corresponding time for the finite-difference extrapolation (46.11 s).

To show that the same effect takes place in case of a rough velocity model, we use first a simple two-layer velocity model, similar to the one used by Fowler, Du and Fletcher (2010). The difference between a dispersion-infested result of the classic finite-difference method (second-order in time, fourth-order in space) and a dispersion-free result of the lowrank approximation is clearly visible in Fig. 5. The time step was 2 ms, which corresponded to the approximation rank of 3. In our implementation, the CPU time for finding the lowrank approximation was 2.69 s and the single-processor CPU time for extrapolation for 601 time steps was 19.76 s or 2.48 times slower than the corresponding time for the finite-difference extrapolation (7.97 s). At larger time steps, the finite-difference method in this model becomes unstable, while the lowrank method remains stable but requires a higher rank.

Next, we move to isotropic wave extrapolation in a complex 2D velocity field. Figure 6 shows a portion of the BP velocity model Billette and Brandsberg-Dahl (2004), containing a salt body. The wavefield snapshot (shown in Fig. 7)

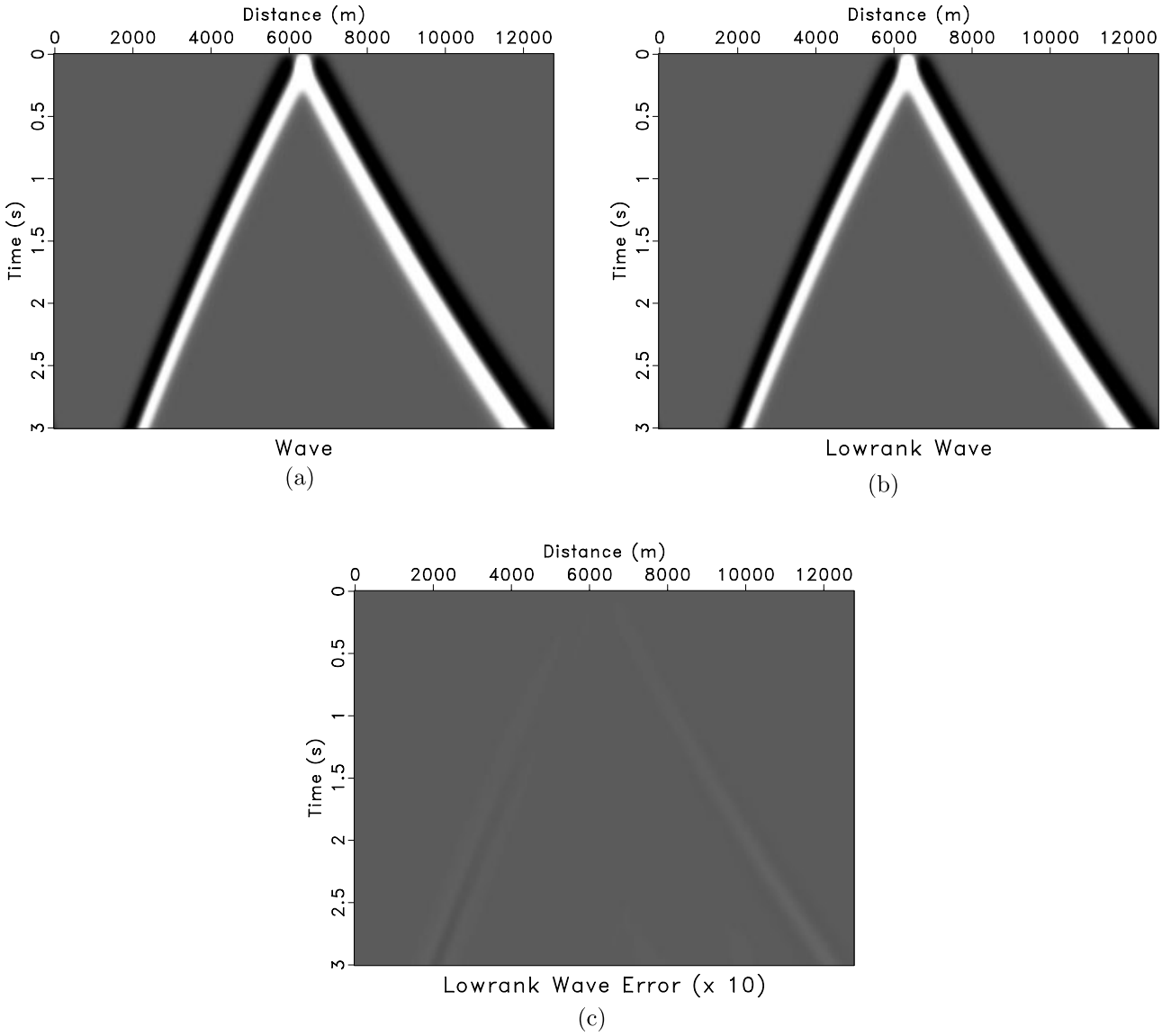

Figure 2 (a) 1D wave extrapolation using the exact extrapolation symbol. (b) 1D wave extrapolation using lowrank approximation. (c) Difference between (a) and (b), with the scale amplified 10 times compared to (a) and (b).

confirms the ability of our method to handle complex models and sharp velocity variations. The lowrank decomposition in this case corresponds to $N = M = 3$, with the expected error of less than 10^{-7}. Increasing the time step size Δt does not break the algorithm but increases the rank of the approximation and correspondingly the number of the required Fourier transforms. For example, increasing Δt from 1 ms to 5 ms leads to $N = M = 5$.

Our next example is isotropic wave extrapolation in a 3D complex velocity field: the SEG/EAGE salt model Aminzadeh, Brac and Kunz 1997 shown in Fig. 8. A dispersion-free wavefield snapshot is shown in Fig. 9. The lowrank decomposition used $N = M = 2$, with the expected error of 10^{-5}.

Finally, we illustrate wave propagation in a complex anisotropic model. The model is a 2007 anisotropic bench-

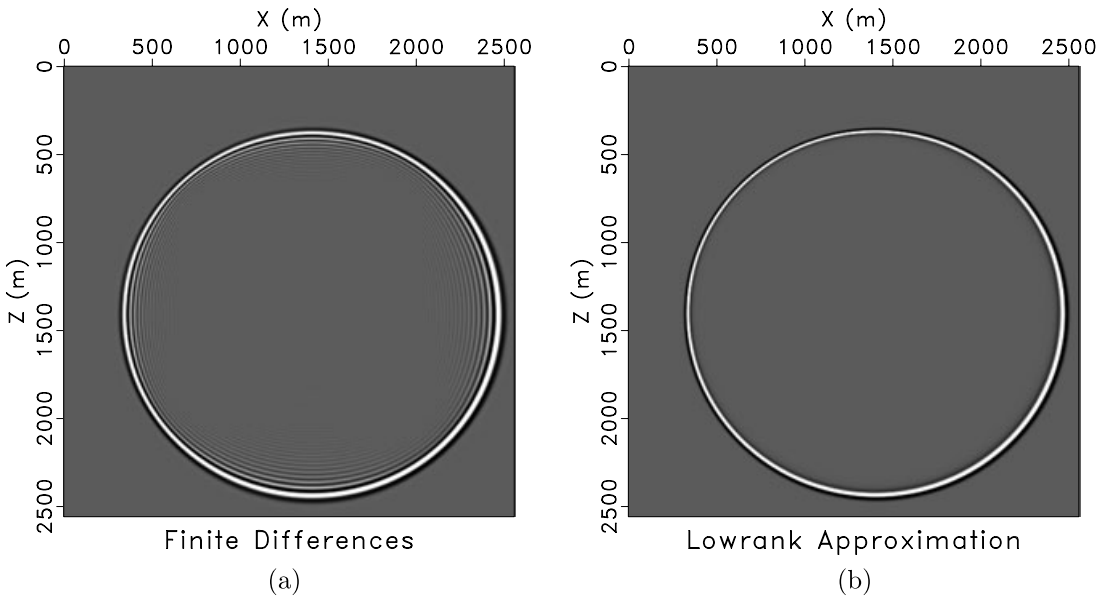

Figure 3 Wavefield snapshot in a smooth velocity model computed using (a) fourth-order finite-difference method and (b) lowrank approximation. The velocity model is $v(x, z) = 550 + 0.00015\,(x - 800)^2 + 0.001\,(z - 500)^2$. The wave source is a point-source Ricker wavelet, located in the middle of the model. The finite-difference result exhibits dispersion artefacts while the result of the lowrank approximation, similarly to that of the FFD method, is dispersion-free.

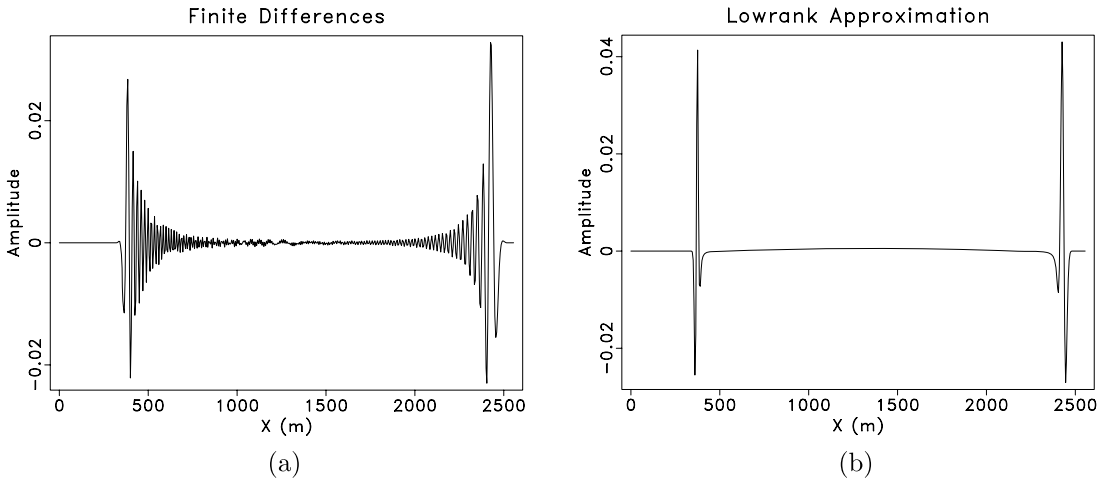

Figure 4 Horizontal slices through wavefield snapshots in Fig. 3.

mark dataset from BP*. It exhibits a strong TTI (tilted transverse isotropy) with a variable tilt of the symmetry axis (Fig. 10). A wavefield snapshot is shown in Fig. 11 . Because of the complexity of the wave propagation patterns, the lowrank decomposition took $N = M = 10$ in this case and required 10 FFTs per time step. In a TTI medium, the phase velocity $V(\mathbf{x}, \mathbf{k})$ from equation (10) can be expressed with the help of the acoustic approximation (Alkhalifah 1998, 2000; Fomel

* The dataset was created by Hemang Shah and is provided at http://software.seg.org/ courtesy of BP Exploration Operation Company Limited.

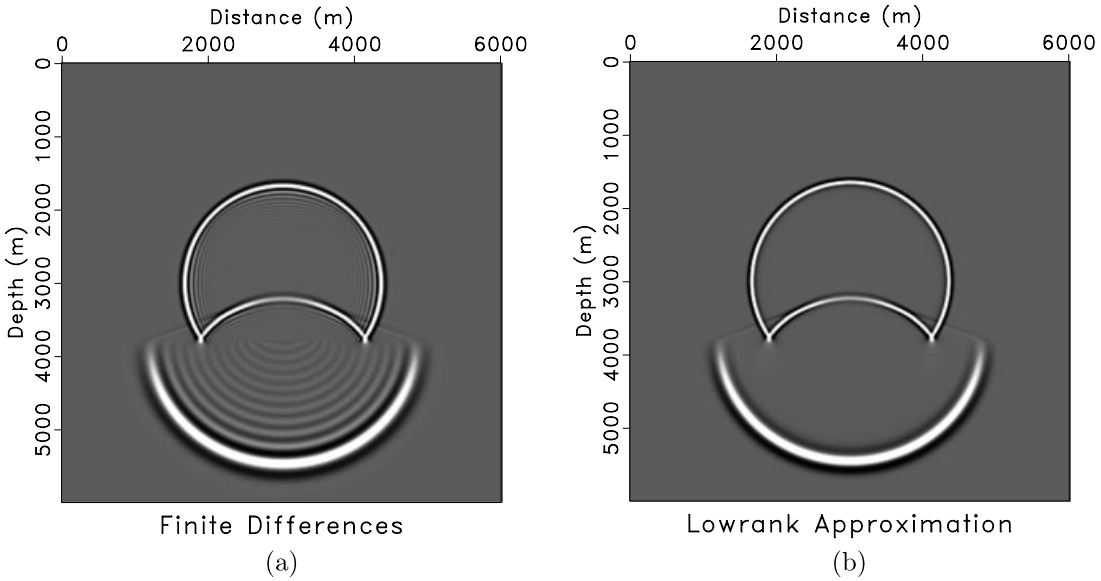

Figure 5 Wavefield snapshot in a simple two-layer velocity model using (a) fourth-order finite-difference method and (b) lowrank approximation. The upper-layer velocity is 1500 m/s, and the bottom-layer velocity is 4500 m/s. The finite-difference result exhibits clearly visible dispersion artefacts while the result of the lowrank approximation is dispersion-free.

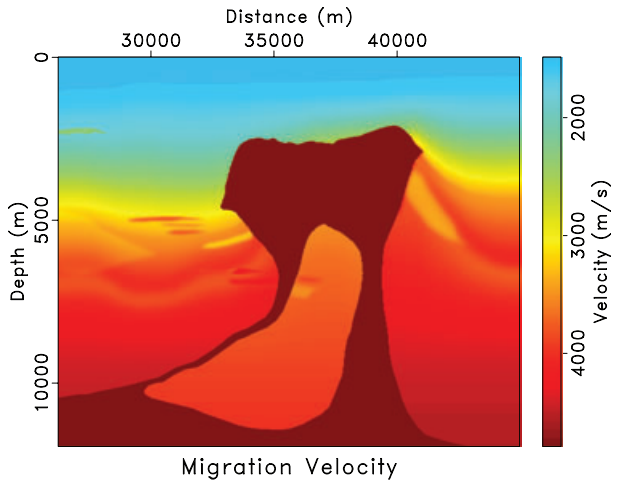

Figure 6 Portion of BP-2004 synthetic isotropic velocity model.

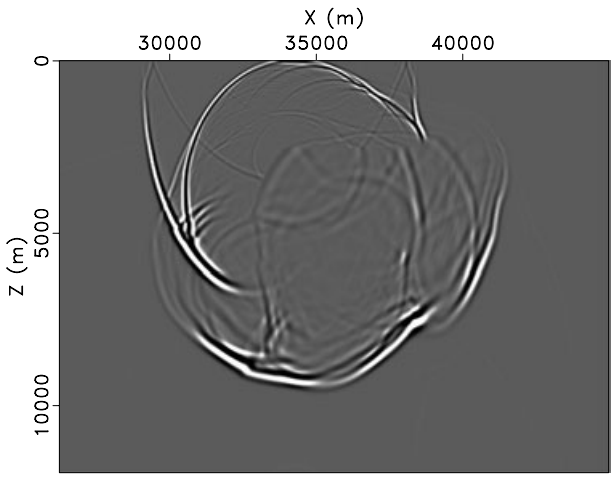

Figure 7 Wavefield snapshot for the velocity model shown in Fig. 6.

2004)

$$V(\mathbf{x},\mathbf{k}) = \sqrt{\frac{1}{2}\left(v_x^2\,\hat{k}_x^2 + v_z^2\,\hat{k}_z^2\right) + \frac{1}{2}\sqrt{\left(v_x^2\,\hat{k}_x^2 + v_z^2\,\hat{k}_z^2\right)^2 - \frac{8\eta}{1+2\eta}v_x^2 v_z^2\,\hat{k}_x^2\,\hat{k}_z^2}}\,, \tag{15}$$

where v_x is the P-wave phase velocity in the symmetry plane, v_z is the P-wave phase velocity in the direction normal to the symmetry plane, η is the anellipticity parameter (Alkhalifah and Tsvankin 1995), and $\hat{k}_x$ and $\hat{k}_z$ stand for the wavenumbers evaluated in a rotated coordinate system aligned with the symmetry axis:

$$\begin{aligned}\hat{k}_x &= k_x\cos\theta + k_z\sin\theta\\ \hat{k}_z &= k_z\cos\theta - k_x\sin\theta\,,\end{aligned} \tag{16}$$

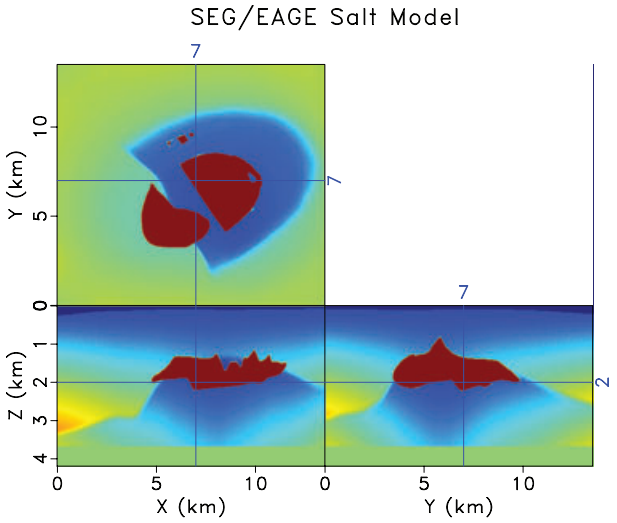

Figure 8 SEG/EAGE 3D salt model.

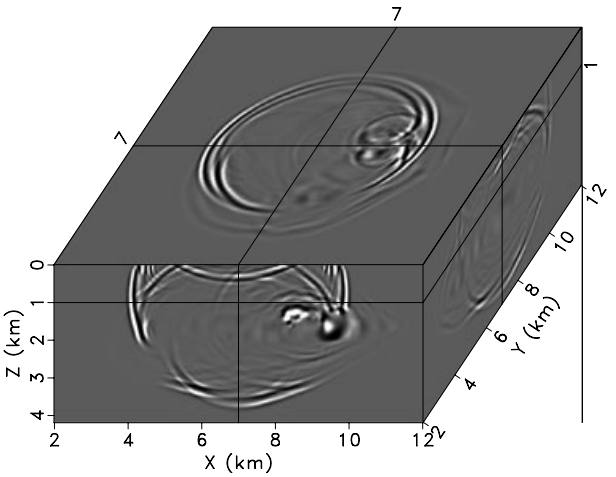

Figure 9 Snapshot of a point-source wavefield propagating in the SEG/EAGE 3D salt model.

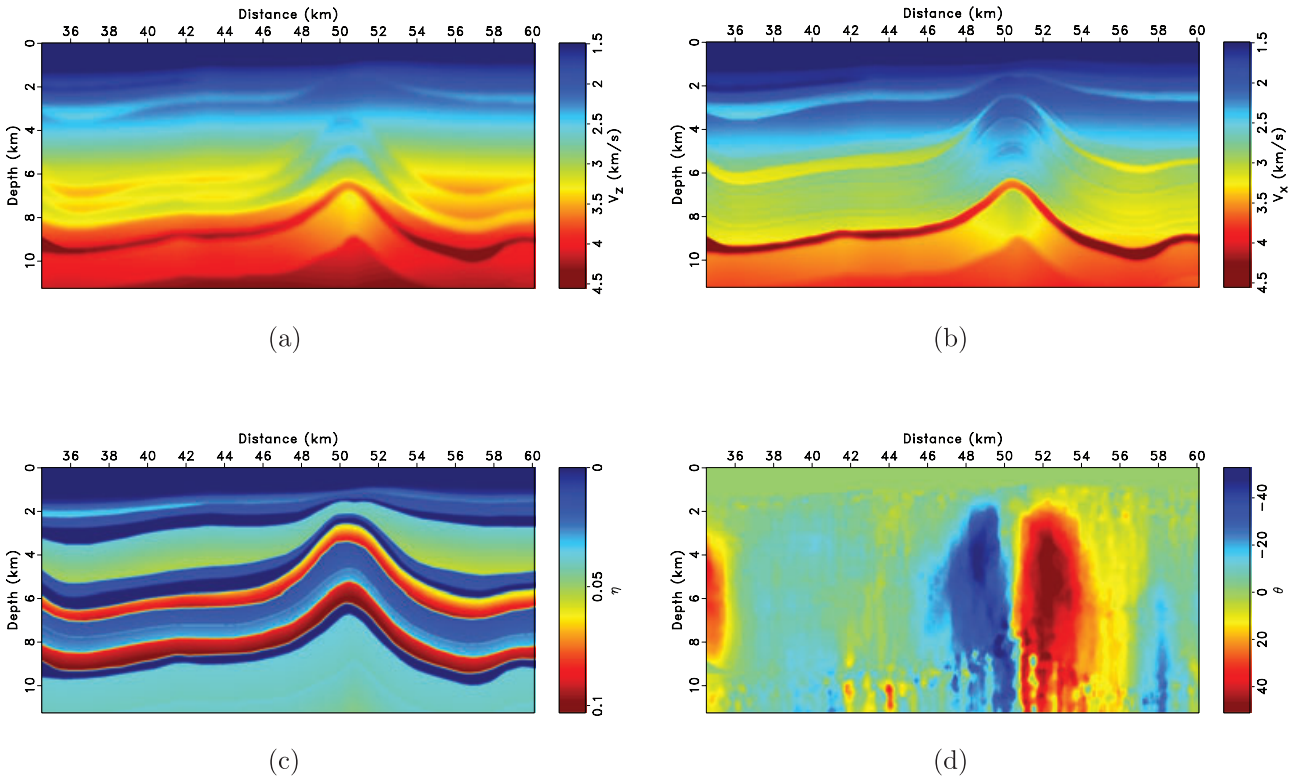

Figure 10 Portion of BP-2007 anisotropic benchmark model. (a) Velocity along the axis of symmetry. (b) Velocity perpendicular to the axis of symmetry. (c) Anellipticity parameter η. (d) Tilt of the symmetry axis.

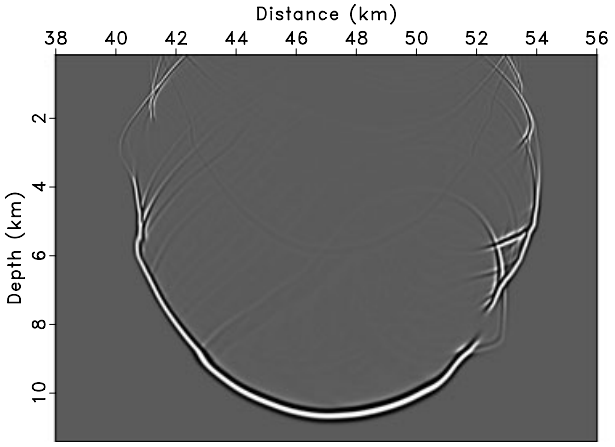

Figure 11 Wavefield snapshot for the velocity model shown in Fig. 10.

where θ is the tilt angle measured with respect to the horizontal.

CONCLUSIONS

We have presented a novel algorithm for wave extrapolation in heterogeneous and anisotropic media. The algorithm is based on a lowrank approximation of the extrapolation symbol. It reduces the cost of extrapolation to that of a small number of FFT operations per time step, which correspond to the approximation rank. The algorithm has a high, spectral accuracy. In that sense, it is comparable with a number of other recently proposed FFT-based methods. Its advantage is a direct control on the accuracy-efficiency trade-off by controlling the rank of the approximation and the corresponding approximation error. We propose to incorporate the lowrank extrapolation algorithm in seismic imaging by reverse-time migration.

ACKNOWLEDGEMENTS

We thank KAUST for partial financial support; Tariq Alkhalifah, Björn Engquist, Laurent Demanet and Paul Fowler for useful discussions; and BP for releasing benchmark synthetic models.

The reproducible computational examples in this paper use the Madagascar open-source software package http://www.ahay.org/.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

Alkhalifah T. 1998. Acoustic approximations for processing in transversely isotropic media. *Geophysics* **63**, 623–631.

Alkhalifah T. 2000. An acoustic wave equation for anisotropic media. *Geophysics* **65**, 1239–1250.

Alkhalifah T. and Fomel S. 2010. Source-receiver two-way wave extrapolation for prestack exploding-reflector modelling and migration. *80th Annual International Meeting, Society of Exploration Geophicists*, 2950–2955.

Alkhalifah T. and Tsvankin I. 1995. Velocity analysis for transversely isotropic media. *Geophysics* **60**, 1550–1566.

Aminzadeh F., Brac J. and Kunz T. 1997. 3D salt and overthrust models. SEG.

Billette F.J. and Brandsberg-Dahl S. 2004. The 2004 BP velocity benchmark. *67th Annual EAGE Meeting*, EAGE, Expanded Abstracts, B305.

Chu C. and Stoffa P. 2008. A pseudospectral-finite difference hybrid approach for large-scale seismic modelling and RTM on parallel computers. *78th Annual International Meeting Society of Exploration Geophicists*, 2087–2091.

Du X., Fletcher R.P. and Fowler P.J. 2010. Pure P-wave propagators versus pseudo acoustic propagators for RTM in VTI media. 72nd Annual EAGE Meeting, EAGE, Expanded Abstracts, Accepted.

Engquist B. and Ying L., 2009. A fast directional algorithm for high frequency acoustic scattering in two dimensions. *Communications Mathematical Sciences* **7**, 327–345.

Etgen J., 1986. High order finite-difference reverse time migration with the two way nonreflecting wave equation. *SEP-48. Stanford Exploration Project*, 133–146.

Etgen J. and Brandsberg-Dahl S. 2009. The pseudo analytical method: Application of pseudo Laplacians to acoustic and acoustic anisotropic wave propagation. *79th Annual International Meeting Society of Exploration Geophysicists*, 2552–2556.

Fomel S. 2003a. Theory of differential offset continuation. *Geophysics* **68**, 718–732

Fomel S. 2003b. Velocity continuation and the anatomy of residual prestack time migration. *Geophysics* **68**, 1650–1661.

Fomel S. 2004. On anelliptic approximations for qP velocities in VTI media. *Geophysical Prospecting* **52**, 247–259.

Fowler P., Du X. and Fletcher R.P. 2010. Recursive integral time extrapolation methods for scalar waves. *80th Annual International Meeting, Society of Exploration Geophysicists*, 3210–3215.

Golub G.H. and Van Loan C.F. 1996. Matrix Computations. John Hopkins.

Goreinov S., Tyrtyshnikov E. and Zamarashkin N. 1997. A theory of pseudoskeleton approximations. *Linear Algebra Applications* **261**, 1–21.

Gu M. and Eisenstat S.C. 1996. Efficient algorithms for computing a strong rank-revealing QR factorization. *SIAM Journal on Scientific Computing* **17**, 848–869.

Johnson W.B. and Lindenstrauss J. 1984. Extensions of Lipschitz mappings into a Hilbert space. In: Conference in modern analysis and probability (New Haven, Connecticut, 1982), American Mathmatical Society, volume **26** *of* Contemporary Mathmatics, 189–206.

Magen A. 2002. Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications. In: Randomization and approximation techniques in computer science. Springer, volume **2483***of* Lecture Notes in Computer Science, 239–253.

Pestana R.C. and Stoffa P.L. 2011. Using the rapid expansion method for accurate time-stepping in modelling and reverse-time migration. *Geophysics* **76**, S177–S185.

Reshef M., Kosloff D. Edwards M. and Hsiung C. 1988. Three-dimensional acoustic modelling by the Fourier method. *Geophysics* **53**, 1175–1183.

Song J. 2001. The optimized expression of a high dimensional function/manifold in a lower dimensional space. *Chinese Scientific Bulletin* **46**, 977–984.

Song X. and Fomel S. 2011. Fourier finite-difference wave propagation. *Geophysics*, accepted.

Song X., Fomel S., Ying L. and Ding T. 2011. Lowrank finite-differences for wave extrapolation. *81st Annual International Meeting, Society of Exploration Geophysicists*, 3372–3376.

Tal-Ezer H., Kosloff D. and Koren Z. 1987. An accurate scheme for seismic forward modelling. *Geophysical Prospecting* **35**, 479–490.

Wards B.D., Margrave G.F. and Lamoureux M.P. 2008. Phase-shift time-stepping for reverse-time migration. *78th Annual International Meeting, Society of Exploration Geophysicists*, 2262–2266.

Zhang Y. and Zhang G. 2009. One-step extrapolation method for reverse time migration. *Geophysics* **74**, A29–A33.

APPENDIX A: LINEAR-TIME ALGORITHM FOR A LOWRANK MATRIX APPROXIMATION

In this appendix, we outline the lowrank matrix approximation algorithm in more detail.

Let N_x be the number of samples both in space and wavenumber. Let us denote the samples in the spatial domain by $\mathbf{x} = \{x_1, \ldots, x_{N_x}\}$ and the ones in the Fourier domain by $\mathbf{k} = \{k_1, \ldots, k_{N_x}\}$. The elements of the interaction matrix $\mathbf{W}$ from equation (11) are then defined as

$$W_{ij} = e^{\,[\phi(x_i, k_j, \Delta t] - x_i \cdot k_j]}, \quad 1 \le i, j \le N_x. \tag{A1}$$

Here we describe an algorithm by Engquist and Ying (2009) that generates, in a time linear with respect to N_x, an approximate factorization of $\mathbf{W}$ of rank r in the following form

$$\mathbf{W} \approx \mathbf{UMV}^* , \tag{A2}$$

where $\mathbf{U}$ consists in r selected columns from $\mathbf{W}$, $\mathbf{M}$ is a matrix of size $r \times r$ and $\mathbf{V}^*$ consists in r selected rows from $\mathbf{W}$.

The first question is: which columns of $\mathbf{W}$ shall one pick for the matrix $\mathbf{U}$? It was shown by Gu and Eisenstat (1996) and Goreinov, Tyrtyshnikov and Zamarashkin (1997) that the r-dimensional volume spanned by these columns should be the maximum or close to the maximum among all possible choices of r columns from $\mathbf{W}$. More precisely, suppose $\mathbf{W} = [w_1, \ldots, w_{N_x}]$ is a column partitioning of $\mathbf{W}$. Then one aims to find $\{j_1, \ldots, j_r\}$ such that

$$\{j_1, \ldots, j_r\} = \mathrm{argmin}_{\{j_1, \ldots, j_r\}} \mathrm{vol}_r(w_{j_1}, \ldots, w_{j_r}). \tag{A3}$$

However, finding a set of r columns with almost the maximum r-dimensional volume is a computationally difficult problem due to the following two reasons. First, the length of the vectors N is typically very large for three-dimensional problems, hence manipulating these vectors can be costly. Second, the number of vectors N_x is also large. A exhaustive search over all possible choices of r vectors to find the one with the maximum volume is prohibitive expensive, so one needs to find a more practical approach.

In order to overcome the problem associated with long vectors, the first idea is to project to a lower dimensional space and search for the set of vectors with maximum volume among the projected vectors. However, one needs to ensure that the volume is roughly preserved after the projection so that the set of vectors with the maximum projected volume also has a near-maximum volume in the original space. One of the most celebrated theorems in high dimensional geometry and probability is the following Johnson-Lindenstrauss lemma (Johnson and Lindenstrauss 1984).

Theorem 1. *Let $v_1, \ldots, v_N$ be a set of N vectors in R^d. Let T be a randomly generated subspace of dimension $t = O(\log N/\varepsilon^2)$ and use P_T to denote the orthogonal projection onto T. Then with high probability,*

$$(1-\varepsilon)\|v_i - v_j\| \le \sqrt{\frac{d}{t}}\|P_T v_i - P_T v_j\| \le (1+\varepsilon)\|v_i - v_j\|$$

for $1 \le i, j \le N$.

This theorem essentially says that projecting to a subspace of dimension $O(\log N)$ preserves the pairwise distance between N arbitrary vectors. There is an immediate generalization of this theorem due to Magen (2002), formulated slightly differently for our purpose.

Theorem 2. *Let $v_1, \ldots, v_N$ be a set of N vectors in R^d. Let T be a randomly generated subspace of dimension $t = O(r^3 \log N/\varepsilon^2)$and use P_T to denote the orthogonal projection onto T. Then with high probability,*

$$(1-\varepsilon)\cdot \mathrm{vol}_r(v_{i_1}, \ldots, v_{i_r}) \le \left(\frac{d}{t}\right)^{r/2} \mathrm{vol}_r(P_T v_{i_1}, \ldots, P_T v_{i_r})$$
$$\le (1+\varepsilon)\cdot \mathrm{vol}_r(v_{i_1}, \ldots, v_{i_r})$$

for any $\{i_1, \ldots, i_r\} \subset \{1, \ldots, N\}$.

The main step of the proof is to bound the singular values of a random matrix between $(1-\varepsilon)^{1/r}$ and $(1+\varepsilon)^{1/r}$ (after a uniform scaling) and this ensures that the r-dimensional volume is preserved within a factor of $(1-\varepsilon)$ and $(1+\varepsilon)$. In order to obtain this bound on the singular values, we need t to be $O(r^3 \log N)$. However, bounding the singular values is only one way to bound the volume, hence it is possible to improve the dependence of t on r. In fact, in practice, we observe that t only needs to scale like $O(r \log N)$.

Given a generic subspace T of dimension t, computing the projections $P_T w_1, \ldots, P_T w_N$ takes $O(tN^2)$ steps. Recall that our goal is to find an algorithm with linear complexity, hence this is still too costly. In order to reduce the cost of the random projection, the second idea of our approach is to randomly choose t coordinates and then project (or restrict) each vector only to these coordinates. This is a projection with much less randomness but one that is much more efficient to apply. Computationally, this is equivalent to restricting $\mathbf{W}$ to t randomly selected rows. We do not yet have a theorem regarding the volume for this projection. However, it preserves the r-dimensional volume very well for the matrix $\mathbf{W}$ and this is in fact due to the oscillatory nature of the columns of $\mathbf{W}$. We denote the resulting vectors by $\{\widetilde{\mathbf{w}}_1, \ldots, \widetilde{\mathbf{w}}_{N_x}\}$.

The next task is to find a set of columns $\{j_1, \ldots, j_r\}$ so that the volume $\mathrm{vol}_r(\widetilde{\mathbf{w}}_{j_1}, \ldots, \widetilde{\mathbf{w}}_{j_r})$ is nearly maximum. As we mentioned earlier, an exhaustive search is too costly. To overcome this, the third idea is to use the following pivoted QR algorithm (or pivoted Gram-Schmidt process) to find the r columns.

1: **for** $s = 1, \ldots, r$ **do**

2: find j_s among $\{1, \ldots, N\} \backslash j_1, \ldots, j_{s-1}$ such that $\widetilde{\mathbf{w}}_{j_s}$ has the largest norm

3: orthogonalize the vectors $\widetilde{\mathbf{w}}_j$ for $j \in \{1, \ldots, N\} \backslash j_1, \ldots, j_s$ with $\widetilde{\mathbf{w}}_{j_s}$ and update them

4: **end for**

5: $\{j_1, \ldots, j_r\}$ is the column set required

Once the column set is found, we set $\mathbf{U} = \left[\mathbf{w}_{j_1}, \ldots, \mathbf{w}_{j_r}\right]$.

In order to identify $\mathbf{V}^*$, one needs to find a set of r rows of $\mathbf{W}$ that has an almost maximum volume. To do that, we repeat the same steps now to $\mathbf{W}^*$. More precisely, let

$$\mathbf{W} = \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_{N_x} \end{bmatrix} \tag{A4}$$

be the row partitioning the matrix $\mathbf{W}$. The algorithm takes the following steps:

1: select uniform randomly a set of t columns and obtain an $N_x \times t$ tall matrix

2: perform pivoted QR algorithm on the rows of this tall matrix and denote the first r rows selected by $\{i_1, \ldots, i_r\}$

3: the matrix $\mathbf{V}^*$ is

$$\mathbf{V}^* = \begin{bmatrix} \mathbf{m}_{i_1} \\ \vdots \\ \mathbf{m}_{i_r} \end{bmatrix}. \tag{A5}$$

Once both $\mathbf{U}$ and $\mathbf{V}^*$ are identified, the last task is to compute the $r \times r$ matrix $\mathbf{M}$ for $\mathbf{W} \approx \mathbf{U}\mathbf{M}\mathbf{V}^*$. Minimizing

$$\min_{\mathbf{M}} \|\mathbf{W} - \mathbf{U}\mathbf{M}\mathbf{V}^*\|_F \tag{A6}$$

yields $\mathbf{M} = (\mathbf{U})^\dagger \mathbf{W} (\mathbf{V}^*)^\dagger$ where $\dagger$ stands for the pseudo-inverse. However, this formula requires taking matrix product with $\mathbf{W}$, which takes $O(t\, N_x^2)$ steps. In order to achieve linear scaling, the fourth idea of our approach is to select randomly a set of t rows A and a set of t columns B and minimize

$$\min_{\mathbf{M}} \|\mathbf{W}(A, B) - \mathbf{U}(A, :)\, \mathbf{M}\, \mathbf{V}(B, :)^*\|_F \,. \tag{A7}$$

The solution for this problem is

$$\mathbf{M} = (\mathbf{U}(A, :))^\dagger\, \mathbf{W}(A, B)\, (\mathbf{V}(B, :)^*)^\dagger \,. \tag{A8}$$

Let us now discuss the overall cost of this algorithm. Random sampling of t rows and t columns of the matrix $\mathbf{W}$ clearly takes $O(t\, N_x)$ steps. Pivoted QR factorization on the projected columns $\{\widetilde{\mathbf{w}}_1, \ldots, \widetilde{\mathbf{w}}_{N_x}\}$ takes $O(t^2\, N_x)$ steps and the cost for the pivoted QR factorization on the projected rows. Finally, performing pseudo-inverses takes $O(t^3)$ steps. Therefore, the overall cost of the algorithm is $O(t\, N_x) + O(t^2\, N_x) + O(t^3) = O(t^2\, N_x)$. As we mentioned earlier, in practice $t = O(r \log N_x)$. Hence, the overall cost is linear in N_x.