We P1 14

Improved FWI Convergence Using Efficient Receiver-side Spatial Preconditioning Employing Ray Theory

G. Yao* (Imperial College London), N. da Silva (Imperial College London), M. Warner (Imperial College London), A. Umpleby (Imperial College London), D. Wu (China University of Petroleum (Beijing))

Summary

Spatial preconditioning can improve the convergence of full-waveform inversion (FWI) significantly. An accurate spatial preconditioning consists of the contribution of both the source and receivers. Prohibited by the unfordable computational cost of directly forming the receiver spatial preconditioning, source-only spatial preconditioning using the energy of the incident source wavefield is typically used to precondition the gradient of FWI. Although this is an efficient means of spatial preconditioning, the quality of the inversion result is still compromised. To improve the quality of spatial preconditioning, we here approximate the receiver spatial preconditioning using ray theory since ray tracing is much faster than numerically solving the two-way wave equation directly. In order to maintain the same time cycle as the inversion without receiver spatial preconditioning, we use an additional compute node to calculate the receiver spatial preconditioning in parallel with other compute nodes used for the usual gradient computation. The effectiveness is demonstrated by applying this technique to the Marmousi model.

Introduction

Full-waveform inversion (FWI) recovers Earth models by optimising an objective function. Restricted by the computational cost, local gradient-descent methods form the natural choice for practical optimisation. In order to speed up convergence, Newton methods can be used for FWI in principal since they utilize the second-order information of the objective function. However, the size and poor-conditioning of the Hessian matrix makes this computationally difficult, and thus it is normally impractical to invert such a matrix. To overcome these problems, *quasi*-Newton methods, e.g. truncated-Newton methods (Nash, 1985) and L-BFGS methods (Liu and Nocedal, 1989), are preferentially in FWI. However, these methods still suffer from increased the computational cost and/or have additional restrictions such as requiring identical data from iteration to iteration.

A robust alternative is to use the diagonal of the Hessian to approximate the full Hessian matrix. The diagonal Hessian is then used to precondition the gradient of FWI. The diagonal Hessian consists of a contribution from the sources and receivers (Burgess and Warner, 2015). To reduce the cost of computing this diagonal to a minimum, it is often approximated using only the energy of the source wavefield. This means that the contribution of the receivers to the diagonal Hessian is ignored. In practice, such source-only spatial preconditioning can improve the convergence of FWI with a steepest-descent method dramatically. Since the source-only spatial preconditioning is effective and is obtained almost for free, it is widely used in FWI. In this paper, we propose a method to increase the effectiveness of spatial conditioning, without a significant increase in cost, by using ray rather than wave theory to approximate the receiver-side contribution to the diagonal of the Hessian.

Theory

The wave equation can be expressed as $\mathbf{Ap} = \mathbf{s}$, where $\mathbf{A}$ denotes the wave equation operator, $\mathbf{p}$ is the wavefield, and $\mathbf{s}$ is the source wavelet. Conventional FWI then minimises an objective function f

$$f = \frac{1}{2}\left(\mathbf{d} - \mathbf{d}_0\right)^{\mathrm{T}}\left(\mathbf{d} - \mathbf{d}_0\right) = \frac{1}{2}\delta\mathbf{d}^{\mathrm{T}}\delta\mathbf{d} \ , \tag{1}$$

where $\mathbf{d}$ denotes the predicted data, $\mathbf{d}_0$ represents the observed data, and $\delta\mathbf{d}$ is the residual. The gradient of the objective function with respective to the model $\mathbf{m}$ can then be written as

$$\frac{\partial f}{\partial \mathbf{m}} = \left(\frac{\partial \mathbf{d}}{\partial \mathbf{m}}\right)^{\mathrm{T}} \delta\mathbf{d} = \mathbf{J}^{\mathrm{T}}\delta\mathbf{d} \ , \tag{2}$$

where $\mathbf{J}$ denotes the Jacobian matrix. Differentiating the wave equation with respective to the model, relating $\mathbf{p}$ and $\mathbf{d}$ using a picking matrix $\mathbf{D}$ so that $\mathbf{d} = \mathbf{Dp}$, and substituting these into Equation 2, gives

$$\mathbf{J} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} = \mathbf{D}\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{D}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} \ . \tag{3}$$

Double differentiating the objective function gives the Hessian matrix $\mathbf{H}$. By assuming $\delta\mathbf{d}$ is small, the Hessian can be approximated as $\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{J}$, and the diagonal elements of the Hessian are then

$$h_{kk} = \left(\mathbf{D}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}_k}\mathbf{p}\right)^{\mathrm{T}}\left(\mathbf{D}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}_k}\mathbf{p}\right) . \tag{4}$$

where k represents the k-th element of the model $\mathbf{m}$. Equation 4 shows that these diagonal elements can be calculated in three steps. First, generate the source wavefield $\mathbf{p}$. Second, scale this by $\partial\mathbf{A}/\partial\mathbf{m}_k$. Finally, forward propagate the scaled wavefield to the receivers. Note that in this last step, the number of wave equation solved is the same as the number of model elements. If we designate the wavefield generated by shot s at the element k by p_{sk}, and the wavefield received at receiver j by r_{skj}, then the k-th element of the diagonal Hessian can be expressed as $h_{kk} = \sum_s \sum_j \sum_i \left(r_{skj}\right)^2$.

If the Green's function from the element k to receiver j is denoted as $G_{kj}(\omega)$ in the frequency domain, then this equation can be written as

$$h_{kk} = \sum_{s}\sum_{i}\left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}_k}\right)^2 p_{sk}^{\dagger}(\omega_i)\, p_{sk}(\omega_i)\sum_{j} G_{kj}^{\dagger}(\omega_i)\, G_{kj}(\omega_i)\,. \tag{5}$$

If the final summation over the Green's functions is removed, then Equation 5 becomes standard source-side spatial preconditioning (Plessix and Mulder, 2004). Instead, here we retain this term to improve the accuracy, but limit the computational cost by using ray theory (Beydoun and Keho, 1987; Červený, 2001) to approximate the Green's function as

$$G_{kj}(\omega) \approx A e^{-j\omega\tau} \tag{6}$$

where τ is the travel time from the receiver to the element, and A is the amplitude calculated through the geometric spreading.

Example

We demonstrate the effectiveness of this receiver-side spatial preconditioning using the Marmousi model, Figure 1a. In this test, 255 shots are generated across the whole model at the surface. A 10-Hz Ricker wavelet is used as the source; a shot record is shown in Figure 1b. The maximum offset was 6.25 km. The initial velocity model, Figure 2a, is a heavily smoothed version of the true model. The inversion consists of 6 blocks, each of which includes 10 iterations. The first 5 blocks invert to a maximum frequency of 3, 4, 6, 8, 10 Hz respectively and the last block uses the full bandwidth.

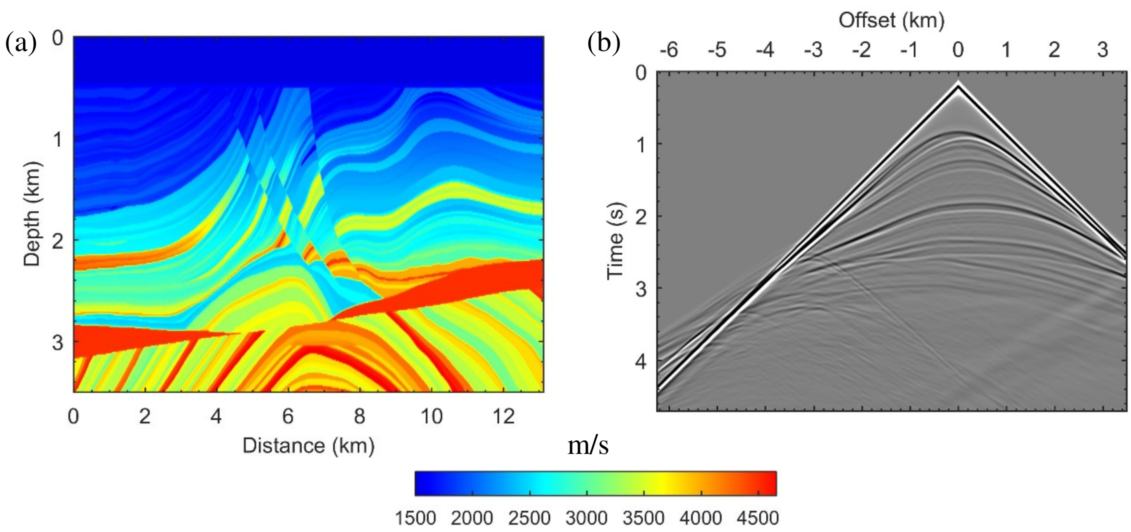


Figure 1 *(a) The Marmousi model, which is discretised into 281 by 1051 cells. (b) A shot record, the source of which is at the distance of 9.64 km.*

During each iteration, we perform the ray tracing to compute the receiver Green's function shown in Equation 6. One receiver's rays is shown in Figure 2a. For one shot, the receiver spatial preconditioning is shown in Figure 2b. Multiplying this preconditioning with its corresponding source spatial preconditioning forms the total spatial preconditioning for this shot. The stack of all shots' spatial preconditioning produces the final spatial preconditioning. Figure 2c shows the final source-only spatial preconditioning while Figure 2d shows the final combined source and receiver spatial preconditioning for the first iteration. Since ray tracing is much faster than solving the wave equation directly, the extra cost associated with the computation of the receiver preconditioning is relatively small. In this example, we used one additional compute node to compute the receiver preconditioning in parallel with 16 compute nodes used for the conventional gradient computations so that the total elapsed time did not increase.

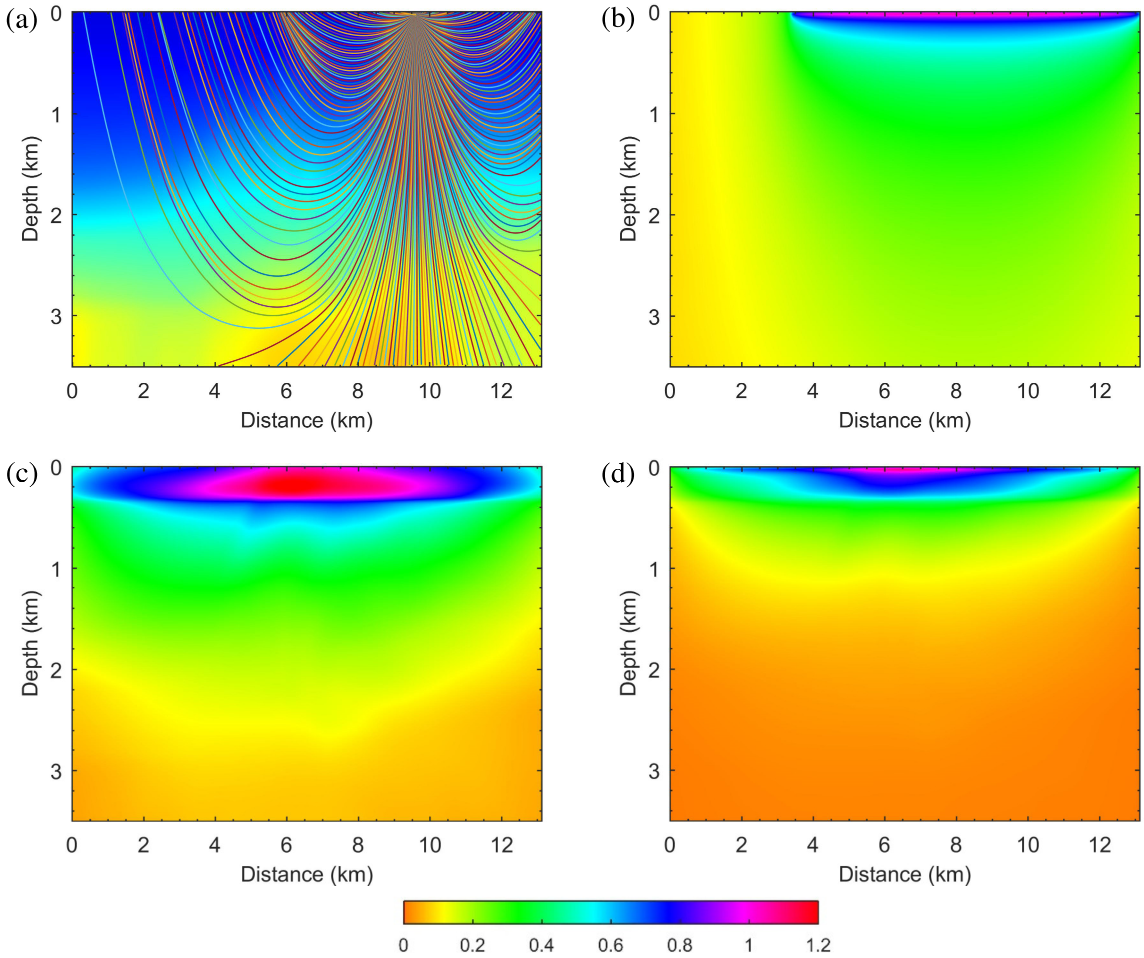


__Figure 2__ Spatial preconditioning in the first iteration. (a) The rays from the receiver at the distance of 9.64 km. The background shows the initial velocity model; its colour scale matches Figure 1. (b) The receiver spatial preconditioning for the shot shown in Figure 1b. (c) The source-only spatial preconditioning. (d) The combined source and receiver spatial preconditioning for the first iteration.

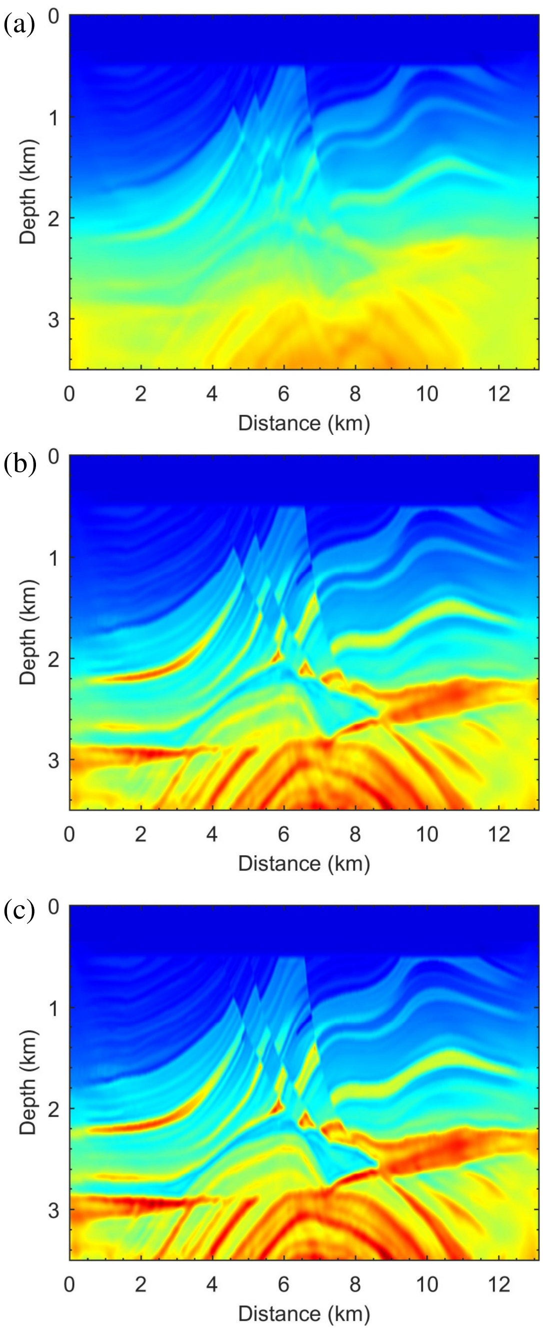
Figure 3 shows the inverted models. By comparison, it clearly shows that spatial preconditioning can improve the quality of inversion results significantly, especially the deep section of the model. The combined source and receiver spatial preconditioning can improve the inversion quality further. These improvements can also be observed through the data residual and the model errors shown in Figure 4.

Conclusions

Spatial preconditioning can improve the convergence of FWI significantly. To improve the accuracy of spatial preconditioning, the receivers' contribution should be considered. Using rays to approximate the receivers' Green's function is an efficient and effective way to compute the receiver spatial preconditioning, and the approach is straightforward to implement and to parallelise in FWI.

Acknowledgements

We thank the sponsors of the FULLWAVE consortium for supporting this work.



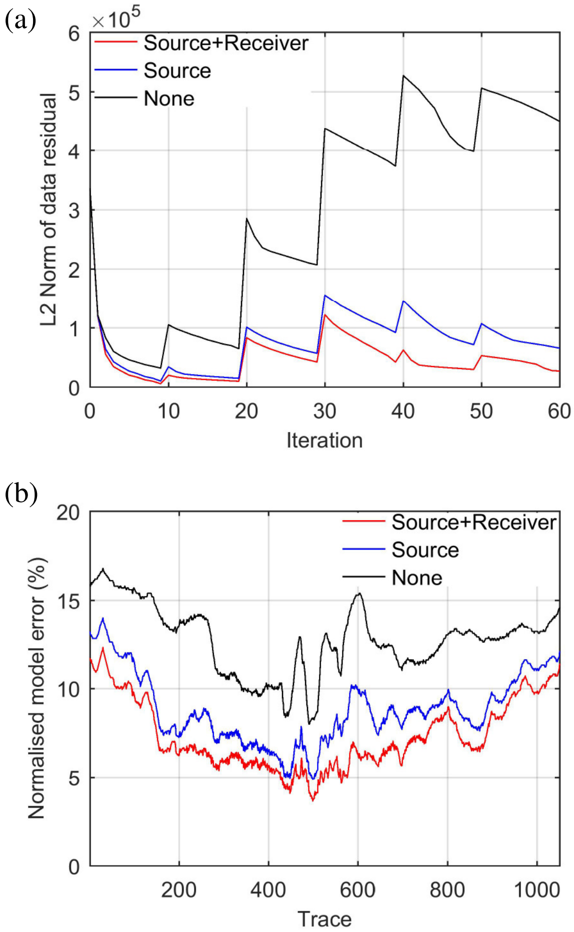


Figure 4 *(a) The data residual, and*
(b) Normalised model errors.

Figure 3 *Inverted models with:*
(a) no spatial pre-conditioning,
(b) source-only spatial preconditioning,
(c) combined source and receiver preconditioning. The colour scale matches that in Figure 1.

References

Beydoun, W.B. and Keho, T.H. [1987] The paraxial ray method. *Geophysics*, **52**, 1639-1653.

Burgess, T. and Warner, M. [2015] Preconditioning FWI with approximate receiver Green's functions. *85th SEG Annual meeting*, Extended Abstracts, 1116-1121.

Červený, V. [2001] *Seismic ray theory*. Cambridge University Press, London.

Liu, D.C. and Nocedal, J. [1989] On the limited memory BFGS method for large scale optimization, *Mathematical Programming*, **45**(1), 503-528.

Nash, S.G. [1985] Preconditioning of Truncated-Newton Methods. *SIAM Journal on Scientific and Statistical Computing*, **6**(3), 599-616.

Plessix, R.-E. and Mulder, W.A. [2004] Frequency-domain finite-difference amplitude-preserving migration. *Geophysical Journal International*, **157**, 975-987.