Crosstalk Reduction in L1-norm-based Simultaneous-source Full Waveform Inversion

W. Jeong, S. Pyun & D.J. Min

SUMMARY

Simultaneous-source technique has been mainly employed for the l2-norm-based full waveform inversion. Considering that real field data are noisy, however, it would be preferable to use robust objective function such as the l1-norm objective function. When we combine the simultaneous-source technique and the l1-norm objective function, there are several problems to be resolved. In case of the l2-norm-based full waveform inversion, the gradient direction of the simultaneous-source inversion has an identical form with that of the individual-source inversion if the crosstalk terms are excluded. In case of the l1-norm-based simultaneous-source inversion, however, the gradient direction is not coincident with that of the individual-source inversion. For this reason, the meaning and the crosstalk reduction of the l1-norm-based simultaneous-source inversion are questionable. In this study, we investigate the gradient direction of the l1-norm-based simultaneous-source inversion and verify the way how the crosstalk noise is suppressed through the statistical approach. Numerical example shows that the l1-norm-based simultaneous-source efficiently reduces the crosstalk and can give reasonable inversion results for noisy data.

Introduction

As full waveform inversion (FWI) has suffered from heavy computational costs, several studies have been devoted to improving the efficiency of FWI. One of them is to assemble recorded shot gathers into several shot gathers or one super-shot gather and to use the assembled shot gathers simultaneously for inversion (Krebs et al., 2009; Ben-Hadj-Ali et al., 2011), which is called 'simultaneous-source method'. This simultaneous-source method has been popularly applied in both seismic migration and FWI, based on the l_2-norm. Although the simultaneous-source method generates crosstalk noise, it can be reduced by the phase encoding function as the iteration goes on (Schuster et al., 2011).

On the other hand, for noisy data with severe deviations, the l_2-norm may not be robust, and the l_1-norm can be preferred (Claerbout and Muir, 1973; Brossier et al., 2010). Several studies have already demonstrated the robustness of the l_1-norm for noisy data. Pyun et al. (2009) showed that the l_1-norm flattens the amplitude spectra of the residual wavefield preserving the phase spectra, which minimizes the effect of outliers and makes the l_1-norm objective function relatively insensitive to noise. Son et al. (2012) proposed combining the l_1-norm FWI with the simultaneous-source methods, which is slightly different from the l_2-norm case. Due to the property of the l_1-norm that the super-shot residual is normalized by its absolute value, which is referred to as 'denominator' in this paper, the simultaneous-source l_1-norm FWI (SS-1) is not equivalent to the individual-source l_1-norm FWI even though the crosstalk is excluded. Although numerical examples showed that the crosstalk did not appear in the final inversion results (Son et al., 2012), crosstalk reduction in the SS-1 is questionable because of the denominator.

In this study, we provide the theoretical background for the SS-1 and verify the crosstalk reduction. We first investigate the convergence of the gradient direction of the SS-1 through the Monte-Carlo simulation, which is able to give us an evidence for the crosstalk reduction of the SS-1. Then, we examine the signal-to-noise ratio (SNR) of the gradient direction through the stacking iteration introduced by Schuster et al. (2011). We also investigate the computational efficiency of the SS-1 comparing it with other FWI algorithms. The simultaneous-source (SS) and individual-source (IS) methods incorporated with the l_1- and l_2-norm objective functions will be referred to as SS-1, SS-2, IS-1 and IS-2 throughout this paper.

Theory

The objective functions of the SS-1 and SS-2 for a single frequency can be expressed by

$$C_{SS1} = \sum_{j=1}^{nr} \left| \sum_{i=1}^{ns} e_i u_{ij} - \sum_{i=1}^{ns} e_i d_{ij} \right| \tag{1}$$

and

$$C_{SS2} = \sum_{j=1}^{nr} \frac{1}{2} \left(\sum_{i=1}^{ns} e_i u_{ij} - \sum_{i=1}^{ns} e_i d_{ij} \right) \overline{\left(\sum_{i=1}^{ns} e_i u_{ij} - \sum_{i=1}^{ns} e_i d_{ij} \right)}, \tag{2}$$

where the subscripts i and j mean the source and receiver indices, respectively. The abbreviated letters ns and nr indicate the total number of sources and receivers, respectively. u_{ij} and d_{ij} denote the modelled and recorded data, respectively. e_i is the encoding function defined as $e_i = \exp(l\phi_i)$, where $l = \sqrt{-1}$ (Ben-Hadj-Ali et al., 2011). The over-bar indicates the complex conjugate. The gradient directions of the SS-1 and SS-2 can be written as the sum of the standard gradient direction and the crosstalk noise:

$$\frac{\partial C_{SS1}}{\partial m} = \sum_{j=1}^{nr} \frac{1}{\left| r_j^{SS} \right|} \left\{ \sum_{i=1}^{ns} \left(e_i \overline{e_i} \frac{\partial u_{ij}}{\partial m} \overline{r_{ij}} \right) + \sum_{i=1}^{ns} \sum_{\substack{k=1 \\ k \neq i}}^{ns} \left(e_i \overline{e_k} \frac{\partial u_{ij}}{\partial m} \overline{r_{kj}} \right) \right\} \tag{3}$$

and

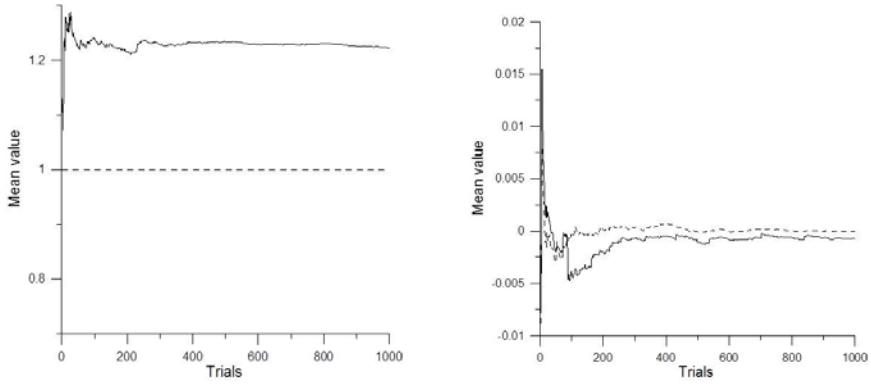
$$\frac{\partial C_{SS2}}{\partial m} = \sum_{j=1}^{nr} \left\{ \sum_{i=1}^{ns} \left(e_i \overline{e_i} \frac{u_{ij}}{\partial m} \overline{r_{ij}} \right) + \sum_{i=1}^{ns} \sum_{\substack{k=1 \\ k \neq i}}^{ns} \left(e_i \overline{e_k} \frac{\partial u_{ij}}{\partial m} \overline{r_{kj}} \right) \right\}, \tag{4}$$

where r_{ij} (= $u_{ij} - d_{ij}$) indicates the residual between modelled and recorded data for the j-th receiver of the i-th shot gather, and r_j^{ss} (= $\sum_{i=1}^{ns} e_i r_{ij}$) denotes the residual at the j-th receiver for a super-shot gather. The second terms on the right-hand side of equations 3 and 4 are regarded as the crosstalk noise that can be minimized by the phase encoding. It is intuitively noted that the gradient direction of the SS-2 (i.e. equation 4) is the same as that of the IS-2 if ignoring the crosstalk term, whereas the gradient direction of the SS-1 is not equivalent to that of the IS-1 due to the denominator. So, to prove the convergence of the SS-1, we compute the expectation value for each gradient direction as follows:

$$E\left(\frac{\partial C_{SS1}}{\partial m}\right) = \sum_{j=1}^{nr}\left\{\sum_{i=1}^{ns} E\left(\frac{e_i \overline{e_i}}{\left|r_j^{SS}\right|}\right)\left(\frac{\partial u_{ij}}{\partial m}\overline{r_{ij}}\right) + \sum_{i=1}^{ns}\sum_{\substack{k=1 \\ k\neq i}}^{ns} E\left(\frac{e_i \overline{e_k}}{\left|r_j^{SS}\right|}\right)\left(\frac{\partial u_{ij}}{\partial m}\overline{r_{kj}}\right)\right\}, \tag{5}$$

$$E\left(\frac{\partial C_{SS2}}{\partial m}\right) = \sum_{j=1}^{nr}\left\{\sum_{i=1}^{ns} E\left(e_i \overline{e_i}\right)\left(\frac{\partial u_{ij}}{\partial m}\overline{r_{ij}}\right) + \sum_{i=1}^{ns}\sum_{\substack{k=1 \\ k\neq i}}^{ns} E\left(e_i \overline{e_k}\right)\left(\frac{\partial u_{ij}}{\partial m}\overline{r_{kj}}\right)\right\}, \tag{6}$$

where $E(\cdot)$ denotes the expectation value. We compute the expectation values at one receiver for 79 sources through the Monte Carlo simulation with 1000 trials for an arbitrarily assumed model. Figure 1 shows the expectation values for the first and second terms on the right-hand sides of equations 5 and 6. As shown in Figure 1a, the expectation value for the first term is 1 in the SS-2 and converges to a certain value in the SS-1. On the other hand, the expectation values for the second terms approach to zero in both methods (Figure 1b). As mentioned above, the first term contributes to the model parameter updates and the second term corresponds to the crosstalk noise which is supposed to be suppressed. Hence, the crosstalk noise can be reduced by iteration for both SS-2 and SS-1. Unlike the SS-2, however, the SS-1 would not exactly converge to the IS-1 because of the denominator (Figure 1a).



(a) (b)

Figure 1 *The expectation values for (a) the first and (b) second terms on the right-hand side of equations 5 and 6, respectively. The black solid and dashed lines denote the SS-1 and SS-2, respectively.*

If we rewrite equation 3 as

$$\frac{\partial C_{SS1}}{\partial m} = \sum_{j=1}^{nr}\left\{\frac{\partial r_j^{SS}}{\partial m}\overline{\left(\frac{r_j^{SS}}{\left|r_j^{SS}\right|}\right)}\right\}, \tag{7}$$

the gradient direction for the SS-1 becomes the same form as the IS-1. The difference is that the residual in the IS-1 is normalized shot by shot by its absolute value while the residual in the SS-1 is normalized all at once by the absolute value of super-shot residual. Although the gradient direction of the SS-1 shown in equation 7 is not equivalent to that of the IS-1, it is reasonable to normalize the super-shot residual by its own absolute value to implement the l_1-norm-based simultaneous-source FWI. Accordingly, we can say that the SS-1 corresponds to the robust form of the SS-2 as the IS-1 does to the IS-2.

Now, we need to investigate whether the crosstalk noise in the gradient direction is actually reduced or not. Following Schuster et al. (2011), we assume that the first terms in equations 5 and 6 are signals

and the crosstalk terms are noise. Then, we compute the SNR curve by using $\mathrm{SNR} = \left\| N \nabla E^{ref} \right\| / \left\| \sum \nabla E_{iter} - N \nabla E^{ref} \right\|$, where $\nabla E^{ref} = \frac{1}{k} \lim_{k \to \infty} \sum_{iter=1}^{k} \nabla E_{iter}$. ∇E^{ref} is calculated by stacking all gradient directions for 5000 iterations. In Figure 2, it is noted that the SS-1 and SS-2 methods are well fitted to the predicted value as shown in Schuster et al. (2011). In addition, the crosstalk reduction of the SS-1 is faster than the SS-2, which may be attributed to the residual normalization that can minimize the crosstalk noise in the amplitude components of the residual.

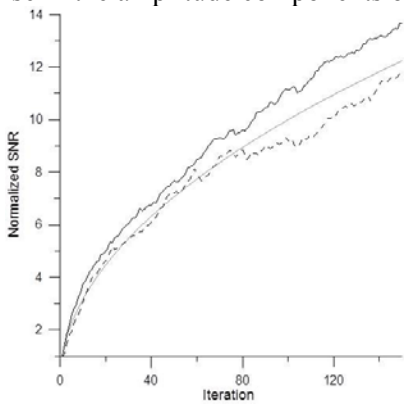
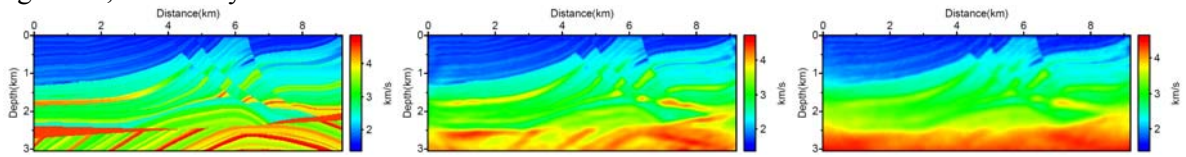


Figure 2 *SNR ratio curves of the gradient direction. The grey solid line indicates predicted SNR and the black solid and dashed lines denote the SNR of the SS-1 and SS-2, respectively.*

Examples

We demonstrated the SS-1 for the modified version of Marmousi-2 model (Martin et al., 2002). Figure 3a shows the P-wave velocity model, which were used as the true model. S-wave velocity model was regenerated with Poisson's ratio of 0.25. The density was assumed to be homogeneous and known. Synthetic data were generated using the finite-difference elastic modelling algorithm in the time domain. The FWI was performed based on the finite-element method in the frequency domain. We assembled the 219 sources to generate a single super-shot. For initial models for FWI, the gradient models were used. Because of limited space, only P-wave velocity results are shown in Figure 3.

Figure 3b shows the inverted results for data with outlier acting like spike noise. In our experiments for the SS-2, deep and complex structures were severely distorted, and the salt structure with high velocity was not recovered due to the outliers. In contrast, although some artefacts are shown in Figure 3b, the SS-1 yields reasonable results even for the outlier-added data.



(a) (b) (c)

Figure 3 *(a) P-wave velocity model of Marmousi-2 model, and SS-1 FWI results for (b) outlier-added data and (c) random noise-added data.*

We also examined the efficiency of the simultaneous-source FWI incorporated with the l_1-norm objective function. To do so, we performed the FWI for a random noise-added data set. To compare the IS-1, IS-2, SS-1 and SS-2 algorithms with each other, we chose a stopping criterion to measure the relative error between the true model and inverted model. The stopping criterion is 12.48% which is the model misfit obtained at the 650th iteration by the IS-2. Each FWI process was iterated until they reach the stopping criterion. The computational speed-up was measured by the forward modelling procedure only considering the number of iterations and the number of sources. The computation time for the factorization of impedance matrix and the other process of the FWI were not considered. We used a Linux cluster machine with 50 Intel Xeon E5630 2.53 GHz CPUs. Figure 3c shows the SS-1 result for the random noise-added data. We only display the result of the SS-1 as a representative due

to the limited space. The computational speed-up is summarized in Table 1. As we expected through the crosstalk reduction, we can see that the SS-1 is faster than the SS-2. From these comparisons, we can conclude that the simultaneous-source FWI preserves the robustness of the l_1-norm objective function for noisy data achieving computational efficiency. The convergence rate of the SS-1 is faster than that of the SS-2.

Table 1 *Computational speed-up of various FWI algorithms. η indicates the computational speed-up*

	error (%)	$\mathbf{N_{iteration}}$	$\mathbf{N_{source}}$	$\boldsymbol{\eta_{theoretical}}$	$\boldsymbol{\eta_{practical}}$
IS-2	12.48	650	219	1	1
IS-1	12.48	597	219	1.09	1.19
SS-2	12.48	2805	1	50.75	54.32
SS-1	12.48	1739	1	81.86	72.59

Conclusions

As a method to improve the robustness of the simultaneous-source FWI, we combined the l_1-norm objective function to the simultaneous-source FWI. Because of the intrinsic property of the l_1-norm, the SS-1 is not mathematically equivalent to the IS-1. However, considering that the l_1-norm is incorporated to the FWI of the super-shot gather, it can be reasonable. To investigate the meaning of the gradient direction of the SS-1, we examined the expectation value of the gradient direction. The expectation value obtained by the Monte Carlo simulation indicates that the SS-1 converges to a certain value which is not equivalent to the IS-1. The SNR curves of the gradient direction for the SS-1 are comparable to the theoretical curves, which also supports that the crosstalk is reduced. The numerical examples showed that the SS-1 properly recover the parameters and is more robust than the SS-2 for data with outliers. Comparing the computational speed-up of the SS-1 with other methods, we noted that the SS-1 shows high performance among the other FWI schemes as shown in crosstalk reduction ratio. From these results, we can conclude that the SS-1 can be efficiently used as a robust method to extract model parameters from noisy data.

Acknowledgements

This work was partly supported by the Energy Efficiency & Resources (No. 2010T100200133) and the Human Resources Development program (No. 20124010203200) of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant by the Korea government Ministry of Knowledge Economy, the "Development of Technology for CO_2 Marine Geological Storage" program funded by the Ministry of Land, Transport and Maritime Affairs (MLTM) of Korea and the Korea CCS R&D Centre (KCRC) grant funded by the Korea government (Ministry of Education, Science and Technology) (No. 2012-0008926).

References

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.

Brossier, R., Operto, S. and Virieux, J. [2010] Which residual norm for robust elastic frequency-domain full waveform inversion?. *Geophysics*, **75**, R37-R46.

Claerbout, J. F. and Muir, F. [1973] Robust modeling with erratic data. *Geophysics*, **38**, 826-844.

Krebs, J., Anderson, J., 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.

Martin, G. S., Marfurt, K. J. and Larsen, S. [2002] Marmousi-2: an updated model for the investigation of AVO in structurally complex areas. *SEG Expanded Absracts*, **21**, 1979-1982.

Pyun, S., Son, W. and Shin, C. [2009] Frequency-domain waveform inversion using an l1-norm objective function. *Exploration Geophysics*, **40**, 227-232.

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**, 1289-1303.