Optimal Green's function estimation for FWI acceleration

Wenyi Hu, Advanced Geophysical Technology Inc.*

Summary

In order to accelerate full waveform inversion (FWI), we developed an efficient optimization algorithm for receiver-side Green's function estimation without any extra forward modeling computation. The optimally estimated receiver-side Green's functions are used to construct the Hessian matrix or an effective preconditioner to enhance the FWI convergence rate. This algorithm, referred to as the optimal Green's function estimation (OGE), produces a series of approximate Green's functions on coarse-to-fine hierarchy lattices. The summation of these series gradually approaches the optimal estimation of the Green's functions with controlled error tolerance. At each resolution level, the algorithm employs a windowed sinc interpolation scheme, serving as an optimal finite impulse filter (FIR) for band-limited signals, to estimate the Green's function at that specific resolution level. The algorithm starts from a low resolution level. At each level, the algorithm estimates the approximation error and uses this error to automatically judge whether the computation should proceed to the next level of resolution. In the conventional FWI, the number of forward modeling required by the popular adjoint method for Hessian computation is $N_S \cdot N_R$ (N_S is the total number of shots and N_R is the number of receivers for each shot), which is computationally prohibitive. With the OGE method, there is no extra forward modeling needed for the Hessian computation. The OGE algorithm is highly automatic and it is independent of shot/receiver geometry. Therefore, the shots and receivers can be distributed highly irregularly. Numerical examples show that this algorithm is very accurate and gives almost identical results to those obtained with explicitly computed Green's functions, while substantially saving the computation expense.

Introduction

Full waveform inversion (FWI) is a method for subsurface geophysical property model reconstruction by minimizing the data misfit between simulated data and recorded data. There are various optimization algorithms for the data misfit minimization. Nonlinear conjugate gradient method (NLCG) is the most popular one but its convergence rate is very slow. For complex velocity model, it takes NLCG hundreds of iterations or even more to converge.

Full Newton and Gauss-Newton are two alternatives that use the Hessian matrix, which includes the 2nd-order derivative information, to accelerate the convergence rate of FWI (Pratt et al., 1998; Hu et al., 2009) greatly. Unfortunately, computation of Hessian or approximate Hessian can be extremely expensive, especially for 3D scenarios. Many efforts have been made to reduce the cost of Hessian computation. Pseudo-Hessian technique was proposed to reduce the computation complexity and it can compensate the geometrical spreading effect to some extent (Shin et al., 2001; Jun et al., 2014). However, many approximations are made in the pseudo-Hessian method, preventing the pseudo-Hessian-based FWI from converging rapidly. Hu et al. (2011) reported two pre-conditioned NLCG (P-NLCG) algorithms to speed up their FWI algorithm, reducing the computation time dramatically. Again, construction of an effective preconditioner requires $N_S \cdot N_R$ additional forward modeling for receiver-side Green's function computation, which is computationally prohibitive for 3D cases. Burgess and Warner (2015) reported a low-cost approach for receiver-side Green's function estimation to construct the diagonal Hessian matrix. However, their approach assumes a homogeneous velocity, leading to a certain level of estimation error. These facts necessitate an efficient and accurate algorithm for receiver-side Green's function computation.

In this work, we developed an algorithm, the so-called optimal Green's function estimation (OGE) for the above mentioned purpose. The idea is that we use the already obtained simulated data to compute the source-side Green's functions and then use these Green's functions to estimate the receiver-side Green's functions through a multi-resolution optimization algorithm with controlled error. First, we assign a set of hidden control lattices in a coarse-to-fine hierarchy order. After that, we estimate the approximate Green's functions in a coarsest lattice first and calculate the approximation error, and then we proceed to the next level of control lattice with higher resolution and compensate the Green's function approximation error in the previous level of lattice. In this iterative manner, we are able to obtain the Green's functions with excitation at the lattice nodes with gradually decreasing estimation errors. Eventually, with these Green's functions whose excitations are distributed regularly on the lattices, the receiver-side Green's functions associated with the irregularly distributed receivers can be evaluated accurately with the same multi-resolution approach. At each resolution level, the Green's function estimation and evaluation are implemented by using a Kaiser windowed sinc interpolation method (Hicks, 2002), a method proved to be very accurate with controlled error. In summary, this approach includes two steps: 1) estimation step - use simulated data of the N_S shots (irregularly distributed) to estimate the Green's functions in a pre-assigned lattices (regularly distributed); 2) evaluation step - use the estimated Green's functions in the lattices (regularly distributed) to evaluate the receiver-side Green's functions (irregularly distributed). Note both the estimation

Optimal Green's function estimation for FWI acceleration

step and the evaluation step are implemented in an iterative multi-resolution manner.

FWI and Hessian Matrix

FWI using Hessian information, Gauss-Newton or preconditioned NLCG, is an inversion algorithm with fast convergence rate. The objective function is

$$C(\mathbf{m}) = \frac{1}{2}\|S - M\|^2, \tag{1}$$

where S is the simulated data and M is the measurement data, $\mathbf{m}$ is the unknown to be inverted. The update amount of the unknown $\Delta\mathbf{m}$ is obtained by

$$\Delta\mathbf{m} = -\gamma \mathbf{H}_a^{-1}\mathbf{g}, \tag{2}$$

where γ is the step-length, $\mathbf{g}$ is the gradient and $\mathbf{H}_a$ is the approximate Hessian matrix

$$\mathbf{H}_a \approx \mathbf{J}^{\mathbf{T}}\mathbf{J}. \tag{3}$$

In (3), $\mathbf{J}$ is the Jacobian matrix, whose element can be computed using the adjoint method

$$J_{s,r;n} = \partial S_{s,r}/m_n = \int_{\tau_n} G^S(\mathbf{r},\mathbf{r_s})G^R(\mathbf{r},\mathbf{r_r})d\mathbf{r}, \tag{4}$$

where s, r, and n denote the indices of the shot, receiver, and a subsurface grid respectively; $G^S(\mathbf{r},\mathbf{r_s})$ is the source-side Green's function at a subsurface position $\mathbf{r}$, with the impulse source applied at the surface source location $\mathbf{r_s}$; $G^R(\mathbf{r},\mathbf{r_r})$ is the receiver-side Green's function at the subsurface position $\mathbf{r}$, with the impulse source applied at the surface receiver location $\mathbf{r_r}$; and τ_n is the support of the subsurface n^{th} grid. $\mathbf{H}_a$ can be further simplified in the P-NLCG methods, but it always requires receiver-side Green's functions $G^R(\mathbf{r},\mathbf{r_r})$ to provide 2nd-order derivative information. The cost of $G^R(\mathbf{r},\mathbf{r_r})$ computation is extremely high because it requires $N_S \cdot N_R$ forward modeling.

Estimation of Receiver-side Green's Function

As discussed in the previous section, the key point of Hessian, approximate Hessian, or a preconditioner construction is the computation of the receiver-side Green's functions. Because explicit receiver-side Green's function computation is prohibitive, we propose an alternative efficient approach to solve this problem.

As shown in Figure 1 (the map view of a seismic survey), our goal is to estimate the receiver-side Greens' functions with the impulse source applied at receiver locations (blue dots). Although the shots and receivers are uniformly distributed in Figure 1, this is just for display purpose. In reality, shots and receivers are irregularly scattered in the survey region and our algorithm can handle arbitrary source/receiver spatial distribution. We also plotted an underlying Cartesian control lattice in Figure 1. There are N_S shots, N_R receivers, and N_P lattice nodes. Assuming that we know the Green's functions at the nodes of the control lattice $G(\mathbf{r},P_i)$, the source-side Green's function can be represented by

$$G(\mathbf{r},S_i) = \sum_{j=1}^{N_p} c_j^{S_i} G(\mathbf{r},P_j), \tag{5}$$

where i is the shot index, $\mathbf{r}$ is a subsurface location, $c_j^{S_i}$ is the interpolation coefficient associated with j^{th} control lattice node for i^{th} shot. (5) can be written in a matrix form

$$\mathbf{G}(\mathbf{r},\mathbf{S}) = \mathbf{C^S}\cdot\mathbf{G}(\mathbf{r},\mathbf{P}), \tag{6}$$

Similarly, we have receiver-side Green's function

$$\mathbf{G}(\mathbf{r},\mathbf{R}) = \mathbf{C^R}\cdot\mathbf{G}(\mathbf{r},\mathbf{P}), \tag{7}$$

where $\mathbf{C^S}$ and $\mathbf{C^R}$ are the interpolation coefficient matrix associated with the shots and the receivers. Combining (6) and (7), we have

$$\mathbf{G}(\mathbf{r},\mathbf{R}) = \mathbf{C^R}\cdot\mathbf{C^{S^+}}\cdot\mathbf{G}(\mathbf{r},\mathbf{S}), \tag{8}$$

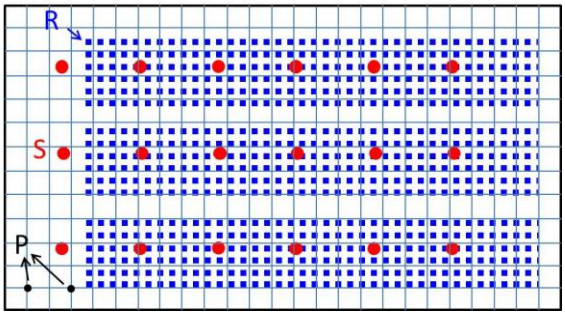

Figure 1: Mapview of shot and receiver distribution and the underlying control lattice.

where + denotes the pseudoinverse. Because $\mathbf{G}(\mathbf{r},\mathbf{S})$ are known from simulated data, we can use (8) to estimate the receiver-side Green's functions $\mathbf{G}(\mathbf{r},\mathbf{R})$ without any extra forward modeling computation. However, although (8) is mathematically correct, we don't have any error control over Green's function estimation. In other words, we don't have any strategy to determine the optimal control lattice resolution. If N_P is too large, the problem is underdetermined, which may result in wrong solutions. On the other hand, if N_P is too small, it becomes an overdetermined problem, which again may lead to an inaccurate $\mathbf{G}(\mathbf{r},\mathbf{R})$. We call this method the single resolution Green's function estimation (SRGE).

Kaiser-Windowed Sinc Interpolation

Before we address the above mentioned issues, we first choose an interpolation scheme to construct the coefficient matrices in (8). The sinc interpolation method is a good choice for band-limited data, especially for frequency domain FWI algorithms. However, there are at least two drawbacks: 1) sinc interpolation method is a global operator; therefore the resulting coefficient matrices are

full, which increases the computational cost; 2) a global interpolation is not necessarily accurate. In our case, because velocity models are spatially varying, short stencil interpolation scheme is more preferable. With these concerns, we apply a spatial window to a sinc function, trying to design a discrete function with finite spatial support to approximate a discrete delta function. We follow the method based on FIR filter design technique proposed by Hicks (2002). The interpolation function is given by

$$f(x) = w(x)\,\mathrm{sinc}(x), \tag{9}$$

$$\mathrm{sinc}(x) \equiv \sin(\pi x)/\pi x, \tag{10}$$

$$w(x) = \frac{I_0\left(\beta\sqrt{1-(x/r)^2}\right)}{I_0(\beta)}, \quad -r \le x \le r \tag{11}$$

where r is the window half-width, I_0 is the zero-order modified Bessel function of the first kind, and β is a parameter adjusting the shape of the Kaiser window. In this work, we choose r=2. With (9) to (11), we are ready to construct the sparse coefficient matrices in (8) and solve for **G**(**r**,**R**) with a predetermined control lattice.

Optimal Green's Function Estimation

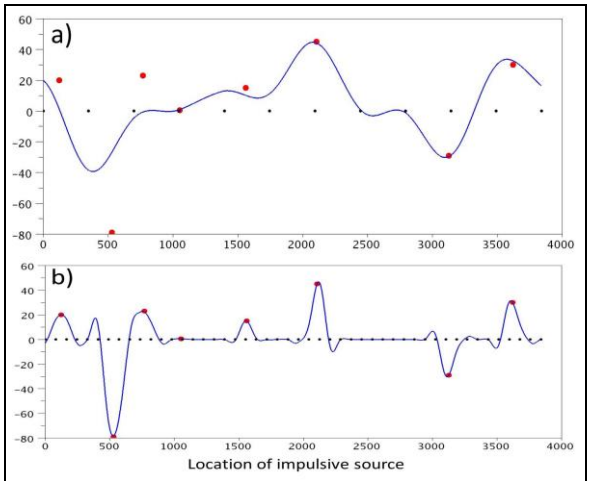

Figure 2: 2D results of single-resolution Green's funciton Estimation (SRGE). Red dots denote the input data (y axis) (i.e., the real part of Green's function at a specific subsurface location) and corresponding source locations (x axis); black dots denote the control lattice node locations; and blue curve represents the estimation result. a) using 12-node control lattice; b) using 48-node control lattice.

The single-resolution Green's function estimation (SRGE) method described by (8) can be recast as a linear inverse problem with the known input data **G**(**r**,**S**) and the unknowns **G**(**r**,**P**), whose objective function is

$$C(\mathbf{G}(\mathbf{r},\mathbf{P})) = \left\|\mathbf{C}^{\mathbf{S}}\mathbf{G}(\mathbf{r},\mathbf{P}) - \mathbf{G}(\mathbf{r},\mathbf{S})\right\|^2 \tag{12}$$

when $N_P >> N_S$, with a predetermined control lattice and a specific interpolation scheme, we can obtain a solution by minimizing the objective function in (12). However, as analyzed earlier, this solution might not be satisfactory even if it fits the data perfectly. On the other hand, if N_P is comparable to N_S or smaller than N_S, the objective function might not be minimized successfully.

Figure 2 shows a simple 2D numerical example (shots are distributed along a 2D line) to illustrate the above mentioned problem. The red dots in Figure 2 give the 7 input data values (red dots' y coordinates, here are the real parts of **G**(**r**,**S**) at a subsurface location **r**) and the corresponding 7 impulse source locations (i.e., red dots' x coordinates, the location where the impulsive source is applied to produce the input data); the black dots indicate the locations of the control lattice nodes; and the blue curve is the estimated result of **G**(**r**,**R**) for arbitrary receiver locations. In Figure 2a), apparently, the estimated result does not match the input data (the blue curve does not pass through the red dots). In Figure 2b), although the data misfit is nearly perfect, the result is not satisfying either. For example, if an impulsive source is applied between 2400 m to 2900 m, the estimated **G**(**r**,**R**) values are always 0, which is unlikely to be true. If we use a finer control lattice, this problem becomes more serious.

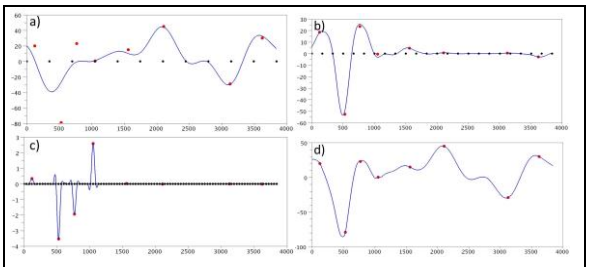

Figure 3: 2D results of optimal Green's funciton Estimation (OGE). a) estimation at 1st resolution level (12-node control lattice); b) compensation at 2nd resolution level (24-node control lattice); c) further compensation at 3rd resolution level (48-node control lattice); d) final estimation result: summation of a)-c).

To circumvent these drawbacks, we developed the optimal Green's function estimation (OGE) algorithm based on multi-resolution analysis. The idea is to introduce a smoothness constraint into the inversion process in an iterative manner. A C^2-continous assumption on the subsurface Green's function volume as a functional of surface impulse source location is reasonable even for highly heterogeneous velocity models. Instead of adding an extra regularization term in the objective function (12), we employ a different strategy – the adaptive progressing regularization technique. The main idea of this strategy is to gradually increase the resolution in the inversion process by introducing a progressively refined control lattice. In

practice, at the first iteration, we use very coarse control lattice to obtain an approximate $\mathbf{G}^1(\mathbf{r},\mathbf{P})$ minimizing

$$C^1 = \left\| \mathbf{C}^{S^1}\mathbf{G}^1(\mathbf{r},\mathbf{P}) - \mathbf{G}(\mathbf{r},\mathbf{S}) \right\|^2 . \quad (13)$$

At each of the following iteration, we refine the control lattice by a factor of 2 to obtain a compensating solution $\mathbf{G}^i(\mathbf{r},\mathbf{P})$, which minimizes

$$C^i = \left\| \mathbf{C}^{S^i}\mathbf{G}^i(\mathbf{r},\mathbf{P}) - \left[\mathbf{G}(\mathbf{r},\mathbf{S}) - \sum_{j=1}^{i-1} \mathbf{C}^{S^j}\mathbf{G}^j(\mathbf{r},\mathbf{P}) \right] \right\|^2 , \quad (14)$$

for $i > 1$. After we obtain all $\mathbf{G}^i(\mathbf{r},\mathbf{P})$ (the estimation step), the receiver-side Green's functions can be evaluated as

$$\mathbf{G}(\mathbf{r},\mathbf{R}) = \sum_{i=1}^{n} \mathbf{C}^{R^i}\mathbf{G}^i(\mathbf{r},\mathbf{P}), \quad (15)$$

which is called the evaluation step.

The numerical example in Figure 2 is re-implemented using the OGE method and the result is shown in Figure 3. The algorithm converges in three iterations. With the increase of the control lattice resolution, the estimation error reduces dramatically. The final estimation result minimizes the data misfit without losing smoothness, which is more preferable in comparison with the SRGE result shown in Figure 2.

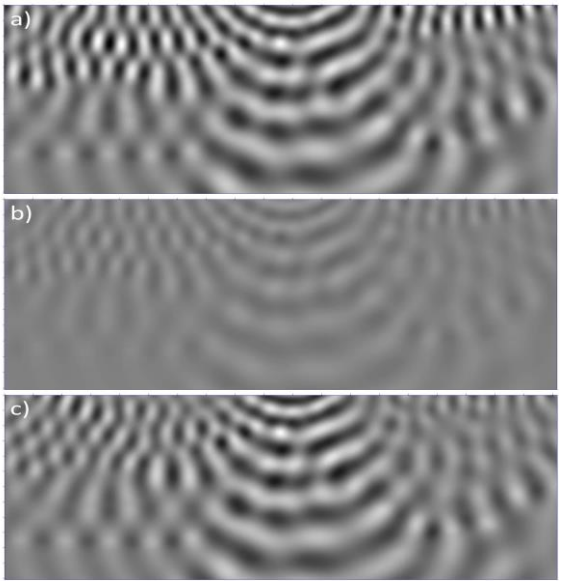

Figure 4: a) Real part of a receiver-side Green's funciton volume for Marmousi model obtained with extra forward modeling computations; b) estimated volume using SRGE with 192-node control lattice; c) esimated volume using OGE with 3 progressive control lattices (48-node, 96-node, and 192-node).

Numerical Results

To evaluate the OGE and SRGE performance and how these algorithms affect the FWI result, we conducted a numerical experiment on the Marmousi model. Figure 4a) shows the real part of a receiver-side Green's function volume explicitly computed using the forward modeling engine. Figure 4b) shows the estimated Green's function volume using SRGE with a 192-node control lattice. Note that the amplitude of the SRGE result is highly underestimated. Figure 4c) is the OGE result using three progressive control lattices with 48-node, 96-node, and 192-node resolution respectively. Combining the OGE and the FWI inversion engine, one is able to accelerate the FWI process significantly by providing the Hessian information without sacrificing the inversion quality, as shown in Figure 5).

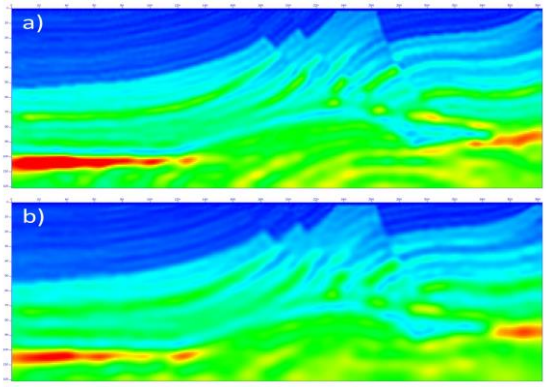

Figure 5: a) Conventional Hessian-based FWI result whose Jacobian matrix is constructed with extra $N_S \cdot N_R$ forward modeling; b) FWI with Jacobian matrix constructed using OGE algorithm without any extra forward modeling.

Conclusions

We developed a novel algorithm to compute receiver-side Green's functions accurately without any additional forward modeling. The estimated Green's functions are used to provide Hessian information to FWI algorithm for convergence rate acceleration. This algorithm, the so-called optimal Green's function estimation (OGE) saves $N_S \cdot N_R$ forward modeling computations compared with the conventional explicit method, making Hessian-based FWI a practical method for production. The OGE is an adaptive inversion algorithm implemented in an iterative manner. At each iteration, the spatial resolution of interpolation is determined by the algorithm automatically to achieve a nearly optimal balance between the data misfit and the spatial smoothness. With iteration proceeds, the data misfit is reduced dramatically and the resolution of the lattice increases gradually. With this strategy, the estimated Green's function is highly accurate and robust without *a priori* information and user interaction. Furthermore, the estimation algorithm employed at each resolution level is based on a windowed sinc interpolation, which ensures the accuracy of the band-limited data interpolation.

References

Burgess, T. and Warner, M., 2015, December. Preconditioning FWI with Approximate Receiver Green's Functions. In 2015 SEG Annual Meeting. Society of Exploration Geophysicists.

Hicks, G. J., 2002, Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions. Geophysics, 67(1), 156-165.

Hu, W., A. Abubakar, and T. M. Habashy, 2009, Simultaneous multifrequency inversion of full-waveform seismic data. Geophysics, 74(2), R1-R14.

Hu, W., A. Abubakar, T. M. Habashy, and J. Liu, 2011, Preconditioned nonlinear conjugate gradient method for frequency domain full waveform seismic inversion, Geophysical Prospecting, 59(3), 477-491.

Jun, H., J. Kim, and C. Shin, 2014, Weighted Pseudo-Hessian Matrix for Frequency-Domain Elastic Full Waveform Inversion, SEG Technical Program Exapanded Abstracts 2014: pp. 1211-1215. doi: 10.1190/segam2014-0622.1

Pratt, R. G., C. Shin, and G. J. Hick, 1998, Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. Geophysical Journal International, 133(2), 341-362.

Shin, C., S. Jang, and D. J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory. Geophysical prospecting, 49(5), 592-606.