Exercise 2 of the Lecture "Full waveform inversion"

Gradient calculation using adjoint state method

Yudi Pan and Thomas Bohlen

17 Nov 2020

Goal

In this exercise, we calculate the gradient of the FWI objective function by using the adjoint state method.

Introduction

The full problem of FWI is to find an optimal model v^{opt} and wavefield p^{opt} which minimize the misfit function. Conventionally, the L2-norm is used to define the misfit function J as:

$$J(v,p) = \frac{1}{2}\sum_{s_r,x_r}\int_0^T \|p^{syn}(t,x_r) - p^{obs}(t,x_r)\|^2 dt + \lambda \cdot F(p,v), \tag{1}$$

where p^{syn} and p^{obs} denotes the synthetic and observed pressure waveforms, respectively. s_r and x_r represents the sources and receivers, respectively, and T represents the total recording time. The relationship between the synthetic data p^{obs} and the model v is constrained by the state equation $F(p,v)$ with a factor λ. $F(p,v)$ in our case is the 2D acoustic wave equation with a constant density:

$$F(p,v) = [\frac{\partial^2}{\partial t^2} - v^2(x,y)\bigtriangledown^2]p(x,y,t) \ - s(x,y,t), \tag{2}$$

If $\hat{p}$ is the solution of $F(\hat{p},v) = 0$ by using a model v, then

$$J(v,\hat{p}) = J(v) = \frac{1}{2}\sum_{s_r,x_r}\int_0^T \|\hat{p}^{syn}(t,x_r) - p^{obs}(t,x_r)\|^2 dt. \tag{3}$$

Therefore, the full objective function is reduced as $J(v)$ in equation 3 when perfect physical engines are used. In other words, we only need to find an optimal model v^{opt}

whose corresponding wavefield $\hat{p}$ calculated by solving the wave equation can minimize the misfit function $J(v)$.

Adjoint state method

The adjoint $A^\dagger$ of a linear operator A is defined as:

$$< Aa, b >=< a, A^\dagger b >, \tag{4}$$

where the scalar product is applied over the time and space dimensions, which is the space Ω with a dimension of R^3 in our 2D case. An operator is called self-adjoint if $A^\dagger = A$.
First of all, we want to derive the adjoint equation of the state equation (or prove that the 2D acoustic wave equation with a constant density is self-adjoint).
Let rewrite the wave equation in an easy way that

$$F(p, v) = \partial_{tt}p - v^2 \bigtriangledown^2 p - s = Ap - Bv^2 p - s = 0, \tag{5}$$

Then we need to prove that A and B are self-adjoint operators. Let's make A as an example, by using high-dimensional derivation by part:

$$\begin{aligned} < Ap, p^\dagger >_\Omega &= < \partial_{tt}p, p^\dagger >_\Omega \\ &= < \partial_t p(T), p^\dagger(T) >_X - < \partial_t p(0), p^\dagger(0) >_X - < \partial_t p, \partial_t p^\dagger >_\Omega \\ &= < \partial_t p(T), p^\dagger(T) >_X - < \partial_t p(0), p^\dagger(0) >_X - < p(T), \partial_t p^\dagger(T) >_X \\ &\quad + < p(0), \partial_t p^\dagger(0) >_X + < p, \partial_{tt} p^\dagger >_\Omega \end{aligned} \tag{6}$$

where X is the 2D spatial space (i.e., $< a, b >_X = \int_{x,y} a\ b\ dxdy$). Due to the initial condition that $\partial_t p \mid_{t=0} = p \mid_{t=0} = 0$:

$$< Ap, p^\dagger >_\Omega =< \partial_t p(T), p^\dagger(T) >_X - < p(T), \partial_t p^\dagger(T) >_X + < p, Ap^\dagger >_\Omega \tag{7}$$

Therefore, A is a self-adjoint operator if $\partial_t p^\dagger \mid_{t=T} = p^\dagger \mid_{t=T} = 0$.
Similarly, we can prove that B (Laplacian operator) is a self-adjoint operator if $\bigtriangledown p \mid_{\hat{X}=\inf} = p \mid_{\hat{X}=\inf} = \bigtriangledown p^\dagger \mid_{\hat{X}=\inf} = p^\dagger \mid_{\hat{X}=\inf} = 0$ where $\hat{X} = \inf$ represents the boundary of the space. Since the radiation and initial boundary conditions have been already included in the wave equation, thus, the 2D acoustic wave equation $F(p, v)$ is self-adjoint when the final boundary condition $\partial_t p^\dagger \mid_{t=T} = p^\dagger \mid_{t=T} = 0$ is satisfied.

Lagrange multiplier

Let's define the Lagrange multiplier from $\Omega \times \Omega^* \times X$ to R (Ω^* is the dual of Ω) by

$$\begin{aligned} \mathcal{L}(p, p^\dagger, v) &= J(v, p) - < p^\dagger, F(p, v) > \\ &= J(v, p) - < p^\dagger, Ap - Bv^2 p - s > \\ &= J(v, p) - < Ap^\dagger - Bv^2 p^\dagger, p > + < p^\dagger, s > . \end{aligned} \tag{8}$$

Notice that p and $p^\dagger$ are independent of model v. If $\hat{p}$ is the simulated wavefield of the model v (i.e., $F(\hat{p}, v) = 0$), then

$$J(v) = \mathcal{L}(\hat{p}, p^\dagger, v) \\ = \frac{1}{2}\|\hat{p} - p^{obs}\|^2 - < Ap^\dagger - Bv^2 p^\dagger, \hat{p} > . \tag{9}$$

According to the optimization theory with equality constraints, $\hat{p}$ is the minimum if $(\hat{p}, p^\dagger)$ is a saddle point of $\mathcal{L}$. Therefore, equating to zero the derivative of $\mathcal{L}$ with respect to p gives the adjoint state equation that

$$Ap^\dagger - Bv^2 p^\dagger = \hat{p}^{syn} - p^{obs}, \text{ s.t. } \partial_t p^\dagger |_{t=T} = p^\dagger |_{t=T} = 0. \tag{10}$$

In order to solve the adjoint state equation (equation 10), we need to back-propagate the residual sources $\delta p = \hat{p}^{syn} - p^{obs}$ at the receivers' points along a reversed time axis $(t' = T - t)$. Therefore, the final boundary condition becomes an initial boundary condition along the reversed time axis that

$$[\frac{\partial^2}{\partial t'^2} - v^2 \bigtriangledown^2] p^\dagger = \delta p(t') \text{ s.t. } \partial_t p^\dagger |_{t'=0} = p^\dagger |_{t'=0} = 0 \text{ and } t' = 0 : T \tag{11}$$

We can see that the adjoint state equation is almost identical to the wave equation and can be solved by using the same code under same stability criteria. Therefore, we could simulate the adjoint wavefield $\hat{p}^\dagger$ by solving equation 11. Recalling that p and $p^\dagger$ are independent of model v, the gradient of the objective function $J(v)$ equals to the gradient of the Lagrange multiplier (equation 9) and can be calculated via:

$$\frac{\partial J(v)}{\partial v} = \frac{\partial \mathcal{L}(\hat{p}, \hat{p}^\dagger, v)}{\partial v} = - < \hat{p}^\dagger, \frac{\partial A\hat{p} - Bv^2\hat{p} - s}{\partial v} >= \frac{2}{v} < \hat{p}^\dagger, \partial_{tt}\hat{p} - s > . \tag{12}$$

Ignoring the influence of source term on the gradient at the source point, we can write the gradient explicitly as

$$\frac{\partial J(v)}{\partial v} = \sum_{s_r} \int_0^T \hat{p}^\dagger(x, y, t' = T - t) \frac{\partial^2 \hat{p}(x, y, t)}{\partial t^2} dt. \tag{13}$$

Therefore, we only need to solve the wave equation *twice* to compute the gradient of the misfit function for *any* number of model parameters N_v.

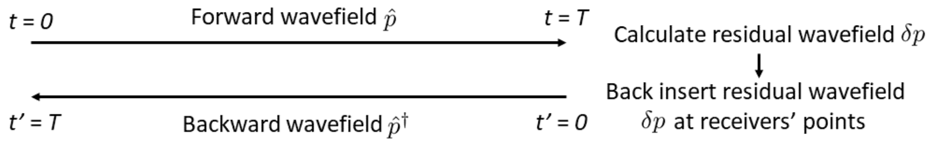
Numerical algorithm

The calculation of gradient for FWI is

1. Compute the forward modeled data $\hat{p}$ at all space-time points.
2. Compute the residual wavefield $\delta p = \hat{p}^{syn} - p^{obs}$ at the receivers' locations.

3. Compute the back-propagated (adjoint) pressure $\hat{p}^{\dagger}$ by using δp as the sources at receivers' locations along a reversed time axis.

4. Compute the gradient $\frac{\partial J(v)}{\partial v}$ (equation 13)

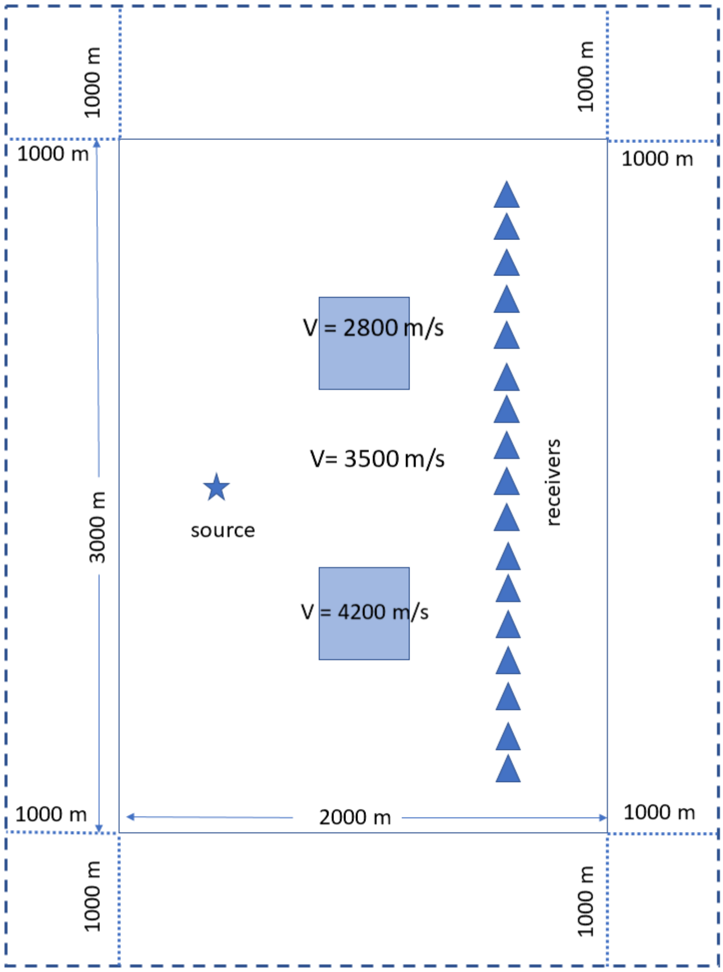


Exercise

1. Prepare a script "gradient.m" to calculate the gradient of FWI. (10 points)
2. Simulate the observed data p^{obs} by using Model 2. (2 points)
3. Calculate the synthetic data $\hat{p}^{syn}$ by using Model 1. Plot the *snapshot* of forward wavefields at 0.2 s, 0.4 s, 0.6 s, and 0.8 s by using the function *imagesc* in matlab. (2 points)
4. Calculate the residual source $\delta p = \hat{p}^{syn} - p^{obs}$ at all receivers positions. Plot out the observed, the synthetic and the residual shot gathers. (2 points)
5. Compute the adjoint wavefield $\hat{p}^{\dagger}$ by backpropagating the time-reversed residual source δp. Plot the *snapshot* of adjoint wavefields at 0.2 s, 0.4 s, 0.6 s, and 0.8 s by using the function *imagesc* in matlab. (2 points)
6. Compute the gradient by using equation 12. Plot the discrete crosscorrelated wavefields $< \hat{p}^{\dagger}, \partial_{tt}\hat{p} >_X$ at 0.2 s, 0.4 s, 0.6 s, and 0.8 s by using the function *imagesc* in matlab. (2 points)
7. Plot the whole gradient by integrating all the discrete crosscorrelated wavefields with respect to time. (2 points)
8. Move the source to (1500,1500), (2000,1500), (3000,1500), (3500,1500) m, respectively. Calculate their corresponding gradients by using the same true and initial models. Stack gradients estimated from all 5 shot gathers and plot the stacked gradient. Compare the stacked gradient to the model residual (true model minus initial model). (8 points)

Model 1 (Initial homogeneous model)
Size of the model : 5000 (y) * 4000 (x) m
Source location: (2500,1500) m
Receivers' location: (1200, 2500), (1250, 2500), (1300, 2500), ..., (3800, 2500) m
Velocity of the model: 3500 m/s
Density of the model: 2500 kg/m^3
Source time function: 10 Hz Ricker wavelet (with 0.3 s delay)
Total recording time: 1.200 s
Grid size: $dx = dz = 25$ m, $dt = 0.004$ s

Model 2 (True crosshole model)
Every parameter is the same as model 1, then adding two rectangular inclusions:
Inclusion 1: v= 2800 m/s (1800 m <= y <= 2200 m, 1800 m <= x <= 2200 m),
Inclusion 2: v= 4200 m/s (2800 m <= y <= 3200 m, 1800 m <= x <= 2200 m),



Reports and scripts

Name your report (in pdf format) and matlab script as "exercise2_YN.pdf" and "gradient_YN.m", respectively, where YN is your name. Send them to yudi.pan@kit.edu and thomas.bohlen@kit.edu before **11.Dec.2020**.