Paper Reading Note

Tche L.

January 20, 2019

Contents

1 Pratt_1997_GJI_Newton methods **4**
1.1 Introduction . . . 4
1.2 Forward . . . 5
1.3 Inversion . . . 5
1.3.1 Gradient method . . . 5
1.3.2 Gradient direction . . . 6
1.3.3 Newton method . . . 6
1.3.4 Gauss-Newton method . . . 7
1.3.5 Full Newton method . . . 7
1.3.6 Exact Hessian . . . 7

2 Pratt_1999_Geophy_Frequency domain inversion **8**
2.1 Introduction and basic principles . . . 8
2.2 Source signature estimation . . . 8

3 Sirgue_2004_Geophy_Temporal frequencies selecting **8**
3.1 Introduction . . . 8
3.2 Waveform inversion . . . 9
3.3 Gradient analysis . . . 10
3.4 The 1D case . . . 10
3.5 Strategy for choosing frequencies . . . 11
3.6 The equivalence between gradient images and migration . . . 11

4 Plessix_2010_SEG_Application to land data set **12**
4.1 Introduction . . . 12
4.2 Full waveform inversion . . . 12

5 Fichtner_2010_EPSL_Full waveform tomography **12**
5.1 Introduction . . . 12
5.2 Seismic anisotropy . . . 14
5.3 Filtering of tomographic images . . . 14
5.3.1 Spherical spline expansion . . . 14
5.3.2 Filtering through spherical convolution . . . 14

6 Tromp_2005_GJI_Adjoint methods **15**
6.1 Introduction . . . 15
6.2 Waveform tomography . . . 16
6.2.1 Topography on internal discontinuities . . . 18
6.3 Adjoint equations . . . 19
6.4 Traveltime tomography . . . 19
6.4.1 Banana-doughnut kernels . . . 20
6.4.2 Misfit kernels . . . 21
6.4.3 Differential traveltime tomography . . . 21
6.5 Amplitude tomography . . . 22
6.5.1 Attenuation . . . 23
6.6 Generalizations . . . 23
6.6.1 Banana-doughnut kernels . . . 24
6.6.2 Misfit kernels . . . 24
6.7 Source inversions . . . 25
6.8 Joint inversions . . . 25

7 Tape_2009_S_Adjoint at SCC **26**
7.1 Introduction . . . 26
7.2 Adjoint tomography . . . 27
7.2.1 Misfit function . . . 27
7.3 Misfit analysis . . . 27
7.4 Earthquake source parameters . . . 28

8 Liu_2006_BSSA_Finite-frequency kernels **28**
8.1 Introduction . . . 28
8.2 Lagrange multiplier method . . . 29
8.3 Absorbing boundaries . . . 31
8.4 Numerical implementation . . . 31

9 Tape_2007_GJI_Adjoint tomography 2D **31**
9.1 Introduction . . . 31
9.2 Inverse problem . . . 33
9.3 Classical tomography . . . 33
9.3.1 Theory . . . 33
9.3.2 Experimental set-up . . . 35
9.4 The gradient . . . 36
9.4.1 Event kernels . . . 36
9.4.2 Misfit kernels . . . 36
9.4.3 Basis function . . . 36
9.5 Optimization . . . 37
9.5.1 Conjugate gradient algorithm . . . 37
9.6 Source, structure and joint inversions . . . 37
9.6.1 Source inversion . . . 37
9.6.2 Joint inversions . . . 38
9.7 Discussion . . . 38
9.7.1 Three kernel types . . . 38

10 Bozdag_2011_GJI_Misfit functions for FWI **38**
10.1 Introduction . . . 38
10.2 Misfit functions and adjoint sources . . . 40
10.2.1 Adjoint kernels . . . 40
10.2.2 Hilbert transform . . . 40
10.2.3 Instantaneous phase misfits . . . 41
10.2.4 Envelope misfits . . . 41
10.2.5 Waveform misfits . . . 42
10.2.6 Traveltime misfits . . . 43
10.2.7 Amplitude misfits . . . 43
10.2.8 Attenuation kernels . . . 43
10.3 Discussion . . . 44

11 Moghaddam_2013_Geophy_Stochastic gradient method **44**
11.1 Introduction . . . 44
11.2 Stochastic optimization . . . 45

12 Louboutin_2017_EAGE_Gradient sampling algorithm **46**
12.1 Introduction . . . 46
12.2 Gradient sampling for FWI . . . 46
12.2.1 Gradient sampling algorithm . . . 46
12.2.2 Implicit time shift . . . 47

13 Schraudolph_2007_AIStat_Stochastic quasi-Newton method **47**
13.1 Introduction . . . 47
13.2 Preliminaries . . . 48
13.2.1 Stochastic gradient descent (SGD) . . . 48
13.2.2 Stochastic meta-descent (SMD) . . . 48
13.2.3 Natural gradient descent (NG) . . . 49
13.3 The (L)BFGS algorithm . . . 49

14 ZhangW_2006_GJI_Traction image method **50**
14.1 Introduction . . . 50
14.2 DRP/opt MacCormack scheme . . . 51
14.3 Compact MacCormack scheme . . . 51
14.4 Interior interface conditions . . . 52

15 ZhangW_2010_Geophy_ADE CFS-PML **52**
15.1 Introduction . . . 52
15.2 Finite difference numerical scheme . . . 53
15.3 CFS-PML using ADE . . . 53
15.4 Free-surface boundary conditions . . . 54
15.5 Optimal parameters . . . 55
15.6 Complete ADE CFS-PML equations . . . 56

Full Waveform Inversion

1 Pratt_1997_GJI_Newton methods[1]

1.1 Introduction

* Wave inversion 1st attempt: Lines and Kelly, 1983 (partial derivatives of the seismogram, wedge-shaped model).

* An important step: Lailly, 1983; Tarantola, 1984 (steepest descent direction for the inverse problem, backpropagate the data residuals and correlate).

* Numerical results of backpropagating methods: Kolb, Collino and Lailly, 1986; Gauthier, Virieux and Tarantola, 1986.

* Extend to elastic and complex problems: Mora, 1987a.

* Apply to the frequency domain with FDFD: Pratt and Worthington, 1990; Pratt, 1990.

* *************************

* Apply to ray theoretical solutions: Beydoun *et al.*, 1989; Lambare *et al.*, 1992.

* Ray-based techniques to real reflection data: Beydoun *et al.*, 1989; Beydoun *et al.*, 1990.

* Outside the ray paradigm to reflection data: Crase *et al.*, 1990.

* Tomography from real cross-borehole data: Zhou *et al.*, 1995 in time-domain; Song, Williamson and Pratt, 1995 & Pratt *et al.*, 1995 in frequency-domain.

* *************************

* Gauss-Newton method with FDFE: Shin, 1988.

* Full Newton method for small problem: Santosa, 1987.

* *************************

* Multiple-souce numerical modeling by FDM: Marfurt, 1984.

* Further developments in FDM: Jo, Shin and Suh, 1996; Stekl and Pratt, 1997.

* The combined FDM/FDI: Tarantola, 1987 (replace functional analysis with matrix algebra).

* *************************

* Matrix algebra of FDM/FDI formalism: Lailly, 1983.

[1] R. G. Pratt, C. Shin and G. J. Hicks, 1997, Geophys. J. Int., Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. Date: 2016/9/3 Sun.

1.2 Forward

Wave equations:

$$\mathbf{M}\ddot{\tilde{\mathbf{u}}}(t) + \mathbf{K}\tilde{\mathbf{u}}(t) = \tilde{\mathbf{f}}(t) \quad \text{or} \quad \text{if viscous, } \mathbf{M}\ddot{\tilde{\mathbf{u}}}(t) + \mathbf{C}\dot{\tilde{\mathbf{u}}} + \mathbf{K}\tilde{\mathbf{u}}(t) = \tilde{\mathbf{f}}(t)$$

where $\mathbf{M}$: mass matrix; $\mathbf{C}$: damping matrix; $\mathbf{K}$: stiffness matrix; $\tilde{\mathbf{f}}$: source terms.

Take FT:

$$\mathbf{Ku}(\omega) + i\omega\mathbf{Cu}(\omega) - \omega^2\mathbf{Mu}(\omega) = \mathbf{f}(\omega)$$

i.e.

$$\mathbf{Su} = \mathbf{f} \quad \text{or} \quad \mathbf{u} = \mathbf{S}^{-1}\mathbf{f}, \quad \text{with } \mathbf{S} = \mathbf{K} - \omega^2\mathbf{M} + i\omega\mathbf{C}$$

when $f_i = \delta_{ij}$ Kronecker delta, $\mathbf{S}^{-1} = [\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \dots, \mathbf{g}^{(l)}]$, where $\mathbf{g}^{(j)}$ approximate the discretized Green's function for an impulse at the jth node, and l is nodal point number.

1.3 Inversion

Residual error: $\delta d_i = u_i - d_i$, $(i = 1, 2, \dots, n)$, where n is the number of receivers.

Minimize the misfit function:

$$\mathbf{E}(\mathbf{p}) = \frac{1}{2}\delta\mathbf{d}^t\delta\mathbf{d}^* = \frac{|\delta\mathbf{d}|^2}{2}$$

where $\mathbf{p}$ is model parameters, and the superscript t and * represent matrix transpose and complex conjugate, respectively.

1.3.1 Gradient method

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - \alpha^{(k)}\nabla_p\mathbf{E}^{(k)}$$

$$\nabla_p\mathbf{E} = \frac{\partial\mathbf{E}}{\partial\mathbf{p}} = \Re\mathrm{e}\{\mathbf{J}^t\delta\mathbf{d}^*\}$$

where $\mathbf{J}$ is the Fréchet derivative matrix and $J_{ij} = \partial u_i/\partial p_j, i = (1, 2, \dots, n), j = (1, 2, \dots, m)$, m is the number of model parameters.

For linear forward problems:

$$\alpha^{(k)} = \frac{|\nabla_p\mathbf{E}|^2}{|\mathbf{J}\nabla_p E|^2}.$$

While for nonlinear forward problems, find $\alpha^{(k)}$ using line-search method.

Augment $\mathbf{J}_{m\times n}$ to $\hat{\mathbf{J}}_{m\times l}$ and $\mathbf{d}_{n\times 1}$ to $\hat{\mathbf{d}}_{l\times 1}$, rewrite:

$$\nabla_p\mathbf{E} = \Re\mathrm{e}\{\hat{\mathbf{J}}^t\delta\hat{\mathbf{d}}^*\}$$

Assuming source is independent of parameter, because of $\mathbf{Su} = \mathbf{f}$:

$$\mathbf{S}\frac{\partial\mathbf{u}}{\partial p_i} = -\frac{\partial\mathbf{S}}{\partial p_i}\mathbf{u} \quad \Rightarrow \quad \frac{\partial\mathbf{u}}{\partial p_i} = \mathbf{S}^{-1}\mathbf{f}^{(i)}, \quad \text{with } \mathbf{f}^{(i)} = -\frac{\partial\mathbf{S}}{\partial p_i}\mathbf{u}$$

where $\mathbf{f}^{(i)}$ is the virtual source term.

1.3.2 Gradient direction

$$\hat{\mathbf{J}} = \Big[\frac{\partial \mathbf{u}}{\partial p_1}, \frac{\partial \mathbf{u}}{\partial p_2}, \ldots, \frac{\partial \mathbf{u}}{\partial p_m}\Big] = \mathbf{S}^{-1}[\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(m)}] \quad \text{or} \quad \hat{\mathbf{J}} = \mathbf{S}^{-1}\mathbf{F}$$

Gradient:

$$\nabla_p \mathbf{E} = \Re\mathfrak{e}\{\hat{\mathbf{J}}^t \delta\hat{\mathbf{d}}^*\} = \Re\mathfrak{e}\{\mathbf{F}^t[\mathbf{S}^{-1}]^t \delta\hat{\mathbf{d}}^*\} = \Re\mathfrak{e}\{\mathbf{F}^t \mathbf{v}\}.$$

Because $\mathbf{S}^{-1}$ is symmetric for seismic source-receiver reciprocal problems,

$$\mathbf{v} = [\mathbf{S}^{-1}]^t \delta\hat{\mathbf{d}}^* = \mathbf{S}^{-1}\delta\hat{\mathbf{d}}^*$$

Another development:

$$\nabla_p \mathbf{E} = \Re\mathfrak{e}\{\hat{\mathbf{J}}^t \delta\hat{\mathbf{d}}^*\} = \Re\mathfrak{e}\{\hat{\mathbf{J}}^{t*} \delta\hat{\mathbf{d}}\}$$

$$\mathbf{w} = \mathbf{v}^* = [\mathbf{S}^{-1}]^{t*}\delta\hat{\mathbf{d}} = [\mathbf{S}^{-1}]^*\delta\hat{\mathbf{d}}$$

where $\mathbf{v}$ and $\mathbf{w}$ are the backpropagated fields.

For the ith component:

$$(\nabla_p \mathbf{E})_i = \Re\mathfrak{e}\{\mathbf{f}^{(i)t}\mathbf{v}\} = \Re\mathfrak{e}\{\mathbf{u}^t\Big[\frac{\partial \mathbf{S}^t}{\partial p_i}\Big]\mathbf{v}\}$$

1.3.3 Newton method

Taylor expansion:

$$\mathbf{E}(\mathbf{p} + \delta\mathbf{p}) = \mathbf{E}(\mathbf{p}) + \delta\mathbf{p}^t \nabla_p \mathbf{E}(\mathbf{p}) + \frac{1}{2}\delta\mathbf{p}^t \mathbf{H}\delta\mathbf{p} + O(|\delta\mathbf{p}|^3)$$

where $\mathbf{H}$ is the $m \times m$ Hessian second-derivative matrix and

$$H_{ij} = \frac{\partial^2 \mathbf{E}(\mathbf{p})}{\partial p_i \partial p_j}, i = (1, 2, \ldots, m), j = (1, 2, \ldots, m)$$

Minimizing with $\delta\mathbf{p}$, take the first-derivative, the solution is

$$\mathbf{H}\delta\mathbf{p} = -\nabla_p \mathbf{E} \quad \text{or} \quad \delta\mathbf{p} = -\mathbf{H}^{-1}\nabla_p \mathbf{E}$$

Newton method for iterative solution:

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - \mathbf{H}^{-1}\nabla_p \mathbf{E}$$

$$H_{ij} = \frac{\partial^2 \mathbf{E}}{\partial p_i \partial p_j} = \Re\mathfrak{e}\{\mathbf{J}^t\mathbf{J}^*\} + \Re\mathfrak{e}\Big\{\Big[\Big(\frac{\partial}{\partial p_1}\mathbf{J}^t\Big)\delta\mathbf{d}^*, \Big(\frac{\partial}{\partial p_2}\mathbf{J}^t\Big)\delta\mathbf{d}^*, \ldots, \Big(\frac{\partial}{\partial p_m}\mathbf{J}^t\Big)\delta\mathbf{d}^*\Big]\Big\} = \mathbf{H}_a + \mathbf{R}$$

where

$$\mathbf{H}_a = \Re\mathfrak{e}\{\mathbf{J}^t\mathbf{J}^*\}$$

$$\mathbf{R} = \Re\mathfrak{e}\Big\{\Big(\frac{\partial}{\partial p_t}\Big)(\delta\mathbf{d}^*, \delta\mathbf{d}^*, \ldots, \delta\mathbf{d}^*)\Big\}$$

1.3.4 Gauss-Newton method

If we neglect the 2nd term $\mathbf{R}$, Gauss-Newton formula:

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - \mathbf{H}_a^{-1}\nabla_p \mathbf{E} \quad \text{and} \quad \delta\mathbf{p} = -\mathbf{H}_a^{-1}\nabla_p \mathbf{E}$$

Apply a damping term of regularization:

$$\delta\mathbf{p} = -(\mathbf{H}_a + \lambda\mathbf{I})^{-1}\nabla_p \mathbf{E} \quad \text{(LM method)}$$

The parameter estimates

$$\delta\hat{\mathbf{p}} = -\mathbf{H}^{\dagger}\nabla_p \mathbf{E} = \gamma\delta\mathbf{p}, \gamma = -\mathbf{H}^{\dagger}\Re\{\mathbf{J}^t\mathbf{J}^*\}$$

where γ is the resolution matrix.

Another interpretation for $\delta\mathbf{p}$:

$$\delta\mathbf{p} = -(\mathbf{K}^t\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^t\delta\mathbf{d}' = -\mathbf{K}^t(\mathbf{K}\mathbf{K}^t + \lambda\mathbf{I})^{-1}\delta\mathbf{d}'$$

where

$$\mathbf{K} = \begin{bmatrix} \Re\{\mathbf{J}\} \\ \Im\{\mathbf{J}\} \end{bmatrix}, \delta\mathbf{d}' = \begin{bmatrix} \Re\{\delta\mathbf{d}\} \\ \Im\{\delta\mathbf{d}\} \end{bmatrix}$$

1.3.5 Full Newton method

Use the exact Hessian matrix:

$$\delta\mathbf{p} = -(\mathbf{H}_a + \mathbf{R})^{-1}\nabla_p \mathbf{E}$$

where

$$\mathbf{R} = \Re\Big\{\Big(\frac{\partial}{\partial\mathbf{p}^t}\mathbf{J}^t\Big)(\delta\mathbf{d}^*, \delta\mathbf{d}^*, \ldots, \delta\mathbf{d}^*)\Big\}$$

$$R_{ij} = \Re\Big\{\Big[\frac{\partial^2\mathbf{u}^t}{\partial p_i\partial p_j}\Big]\delta\hat{\mathbf{d}}^*\Big\}, i = (1, 2, \ldots, m), j = (1, 2, \ldots, m)$$

with the augment of $\mathbf{d}_{n\times 1}$ to $\hat{\mathbf{d}}_{l\times 1}$.

1.3.6 Exact Hessian

From

$$\mathbf{S}\frac{\partial\mathbf{u}}{\partial p_i} = -\frac{\partial\mathbf{S}}{\partial p_i}\mathbf{u}$$

take the derivative of p_j to both sides:

$$\mathbf{S}\frac{\partial^2\mathbf{u}}{\partial p_j\partial p_i} + \Big(\frac{\partial\mathbf{S}}{\partial p_j}\Big)\Big(\frac{\partial\mathbf{u}}{\partial p_i}\Big) = -\Big(\frac{\partial\mathbf{S}}{\partial p_i}\Big)\Big(\frac{\partial\mathbf{u}}{\partial p_j} - \frac{\partial^2\mathbf{S}}{\partial p_j\partial p_i}\mathbf{u}$$

i.e.

$$\mathbf{S}\frac{\partial^2\mathbf{u}}{\partial p_j\partial p_i} = -\mathbf{f}^{(ij)} \quad \text{or} \quad \frac{\partial^2\mathbf{u}}{\partial p_j\partial p_i} = -\mathbf{S}^{-1}\mathbf{f}^{(ij)}$$

where

$$\mathbf{f}^{(ij)} = \Big(\frac{\partial\mathbf{S}}{\partial p_i}\Big)\Big(\frac{\partial\mathbf{u}}{\partial p_j}\Big) + \Big(\frac{\partial\mathbf{S}}{\partial p_j}\Big)\Big(\frac{\partial\mathbf{u}}{\partial p_i}\Big) + \frac{\partial^2\mathbf{S}}{\partial p_j\partial p_i}\mathbf{u}$$

is the 2nd-order virtual source term.

Because of

$$\left[\frac{\partial^2 \mathbf{u}}{\partial p_i \partial p_j}\right]^t = -[\mathbf{f}^{(ij)}]^t[\mathbf{S}^{-1}]^t$$

obtain:

$$R_{ij} = -\Re e\{[\mathbf{f}^{(ij)}]^t \mathbf{v}\}, \mathbf{v} = [\mathbf{S}^{-1}]^t \delta\hat{\mathbf{d}}^* = \mathbf{S}^{-1}\delta\hat{\mathbf{d}}^*$$

2 Pratt_1999_Geophy_Frequency domain inversion[2]

2.1 Introduction and basic principles

Same as the former one (Pratt_1997_GJI_Newton methods), i.e. the gradient method.

2.2 Source signature estimation

Assume source signature is scaled source terms in the modeling,

$$\mathbf{S}\mathbf{u} = s\mathbf{f}$$

where **s** is an unknown complex-value scalar.

Using the misfit function:

$$\mathbf{E} = {}^{1}\!/_{2}\delta\mathbf{d}^t\delta\mathbf{d}^*$$

The minimum misfit is found when

$$s = \frac{\mathbf{u}^t\mathbf{d}^*}{\mathbf{u}^t\mathbf{u}^*}$$

3 Sirgue_2004_Geophy_Temporal frequencies selecting[3]

3.1 Introduction

* Wave inversion implementation: Tarantola, 1986 & Mora, 1987 & Burks *et al.*, 1995 & Shipp and Singh, 2002 in time domain; Pratt and Worthington, 1990 & Liao and McMechan, 1996 in frequency domain.
* *************************
* Time windowing the residuals: Shipp and Singh, 2002 in time domain; Mallick and Frazer, 1987 in frequency domain.
* Low-pass filter the data: Bunks *et al.*,1995.
* *************************
* Single frequency yields finite information of the model: Wu and Toksöz, 1987.
* *************************
* Limited number of frequencies would suffice: Freudenreich and Singh, 2000.

[2]R. Gerhard Pratt, 1999, Geophysics, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. Date: 2016/9/14 Wen.

[3]Laurent Sirgue and R. Gerhard Pratt, 2004, Geophysics, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. Date: 2016/9/23 Fri.

* *************************
* Image stretch (NMO stretch) of prestack depth migration: Gardner *et al.*, 1974.
* Stretch effect compensate lack of low frequencies and improve the spectral content of stacked data: Haldorsen and Farmer, 1989.
* Prestack depth imaging for reflection data: Tarantola, 1986.
* Frequency domain prestack depth migration: Schleicher *et al.*, 1993.
* *************************
* Compute step length of iterative gradient method using linear estimate: Tarantola, 1984a; Mora, 1987.
* Conjugate gradient method: Concus *et al.*, 1976.
* Compute gradient without explicitly partial derivatives of the data: Lailly, 1983; Tarantola, 1987; Pratt and Worthington, 1990; Pratt *et al.*, 1998.
* Wavepath as the adjoint of the Fréchet partial derivative: Woodward, 1992.
* Diffraction tomography: Devaney, 1981; Wu and Toksöz, 1987.
* Linearized inversion in the (ω, k) domain: Clayton and Stolt, 1981; Ikelle *et al.*, 1986.

3.2 Waveform inversion

Constant-density acoustic-wave equation:

$$\Big(\nabla^2 + \frac{\omega^2}{c^2(\mathbf{x})}\Big)\Psi(\mathbf{x}, \mathbf{s}, \omega) = -\delta(\mathbf{x} - \mathbf{s})$$

and the model parameter

$$m(\mathbf{x}) = \frac{1}{c^2(\mathbf{x})}$$

where $\Psi(\mathbf{x}, \mathbf{s}, \omega)$ is the pressure field at the spatial location $\mathbf{x}$ with the source location $\mathbf{s}$.

If ω is implicit, the complex-valued data residuals with source-receiver coordinates $\mathbf{s}$ and $\mathbf{r}$:

$$\Delta\Psi(\mathbf{r}, \mathbf{s}) = \Psi_{calc}(\mathbf{r}, \mathbf{s}) - \Psi_{obs}(\mathbf{r}, \mathbf{s})$$

Minimize the misfit function:

$$E = \frac{1}{2}\sum_s \sum_r \delta\Psi^*(\mathbf{r}, \mathbf{s})\delta\Psi(\mathbf{r}, \mathbf{s})$$

where $*$ denotes complex conjugation. And the descent direction:

$$g(\mathbf{x}) = -\nabla_m E = -\frac{\partial E}{\partial m(\mathbf{x})}$$

The model updated by:

$$m(\mathbf{x})^{l+1} = m(\mathbf{x})^l + \gamma^l g(\mathbf{x})^l$$

Compute gradient by zero-lag correlation of the forward propagated wavefield and the back-propagated wavefield (Pratt *et al.*, 1996, eq.12):

$$g(\mathbf{x}) = -\omega^2 \sum_s \sum_r \Re\mathrm{e}\{P_f^*(\mathbf{x}, \mathbf{s}) P_b(\mathbf{x}, \mathbf{r}, \mathbf{s})\}$$

$$P_f(\mathbf{x},\mathbf{s}) = G_0(\mathbf{x},\mathbf{s}) \quad \text{and} \quad P_b(\mathbf{x},\mathbf{r},\mathbf{s}) = G_0^*(\mathbf{x},\mathbf{r})\Delta\Psi(\mathbf{r},\mathbf{s})$$

where $P_f(\mathbf{x},\mathbf{s})$ and $P_b(\mathbf{x},\mathbf{r},\mathbf{s})$ are the forward propagated wavefield of an unit impulsive point source and the back-propagated wavefield of the data residuals, respectively; $G_0(\mathbf{x},\mathbf{s})$ and $G_0(\mathbf{x},\mathbf{r})$ are the Green's functions for exciting at the source and receiver locations, respectively.

The full expression:

$$g(\mathbf{x}) = -\omega^2 \sum_s \sum_r \Re\mathfrak{e}\{G_0^*(\mathbf{x},\mathbf{s})G_0^*(\mathbf{x},\mathbf{r})\Delta\Psi(\mathbf{r},\mathbf{s})\}$$

Assume ignoring amplitude effects, the homogeneous reference medium with velocity c_0 and the far field, approximate by plane waves:

$$G_0(\mathbf{x},\mathbf{s}) \approx \exp(ik_0\hat{\mathbf{s}}\cdot\mathbf{x}) \quad \text{and} \quad G_0(\mathbf{x},\mathbf{r}) \approx \exp(ik_0\hat{\mathbf{r}}\cdot\mathbf{x})$$

where $k_0 = \omega/c_0$ is the wavenumber, and $\hat{\mathbf{s}}$ and $\hat{\mathbf{r}}$ are unit vectors from source (incident propagation) and receiver (inverse scattering) to scatter, respectively. So that

$$\begin{aligned} g(\mathbf{x}) &= -\omega^2 \sum_s \sum_r \Re\mathfrak{e}\{\exp(-ik_0\hat{\mathbf{s}}\cdot\mathbf{x}) \times \exp(-ik_0\hat{\mathbf{r}}\cdot\mathbf{x})\Delta\Psi(\mathbf{r},\mathbf{s})\} \\ &= -\omega^2 \sum_s \sum_r \Re\mathfrak{e}\{\exp(-ik_0(\hat{\mathbf{s}}+\hat{\mathbf{r}})\cdot\mathbf{x})\Delta\Psi(\mathbf{r},\mathbf{s})\} \end{aligned}$$

Note that this is an inverse Fourier summation.

3.3 Gradient analysis

Through the Born approximation (Miller *et al.*, 1987, eq.8):

$$\Delta\Psi(\mathbf{r},\mathbf{s}) \approx -\omega^2 \int d\mathbf{x} G_0(\mathbf{r},\mathbf{x})G_0(\mathbf{x},\mathbf{s})\delta m(\mathbf{x})$$

where $\delta m(\mathbf{x})$ is the true parameter perturbation. Because of the plane-wave approximations, obtain:

$$\Delta\Psi(\mathbf{r},\mathbf{s}) \approx -\omega^2 \int d\mathbf{x}\delta m(\mathbf{x}) \exp(+ik_0(\hat{\mathbf{s}}+\hat{\mathbf{r}})\cdot\mathbf{x})$$

And rewrite as

$$\Delta\Psi(\mathbf{r},\mathbf{s}) = -\omega^2 \tilde{M}(k_0(\hat{\mathbf{s}}+\hat{\mathbf{r}}))$$

where $\tilde{M}(\mathbf{k})$ is the Fourier transform of $\delta m(\mathbf{x})$.

Thus,

$$g(\mathbf{x}) = \omega^4 \sum_s \sum_r \Re\mathfrak{e}\{\exp(-ik_0(\hat{\mathbf{s}}+\hat{\mathbf{r}})\cdot\mathbf{x})\tilde{M}(k_0(\hat{\mathbf{s}}+\hat{\mathbf{r}}))\}$$

This is an inverse Fourier summation where the weights in the summation are given by the Fourier components of the model. And

$$g(\mathbf{x}) \to \omega^4 \delta m(\mathbf{x})$$

where the gradient will recover a scaled image of the original model.

3.4 The 1D case

For a 1D earth (velocity varies only as a function of depth), the incident and scattering angles are symmetric,

$$k_0\hat{\mathbf{s}} = (k_0\sin\theta, k_0\cos\theta) \quad \text{and} \quad k_0\hat{\mathbf{r}} = (k_0\sin(-\theta), k_0\cos(-\theta)) = (-k_0\sin\theta, k_0\cos\theta)$$

where the angles θ and $-\theta$ are for the source and receiver wave, and

$$\cos\theta = \frac{z}{\sqrt{h^2+z^2}} \quad \text{and} \quad \sin\theta = \frac{h}{\sqrt{h^2+z^2}}$$

in which h is the half offset and z is the depth of the scattering layer. So the wavenumber illumination:

$$k_0(\hat{\mathbf{s}}+\hat{\mathbf{r}}) = (k_x, k_z) = (0, 2k_0\alpha) \quad \text{with} \quad \alpha = \cos\theta = \frac{1}{\sqrt{1+R^2}}$$

where $R = {}^{h}/_{z}$.

3.5 Strategy for choosing frequencies

For an offset range $[0, x_{max}]$ of a 1D thin layer, the vertical wavenumber coverage

$$k_z \in [k_{zmin}, k_{zmax}] = [2k_0\alpha_{min}, 2k_0] \quad \text{with} \quad \alpha_{min} = \frac{1}{\sqrt{1+R_{max}^2}}, \alpha_{max} = 1$$

where $R_{max} = {}^{h_{max}}/_{z}$ and h_{max} is the maximum half offset. Due to $k_0 = {}^{\omega}/_{c_0}$, in terms of frequency,

$$k_{zmin} = 4\pi f\alpha_{min}/c_0 \quad \text{and} \quad k_{zmax} = 4\pi f/c_0$$

Define the wavenumber coverage and the wavenumber bandwidth

$$\Delta k_z \triangleq |k_{zmax} - k_{zmin}| = 4\pi(1-\alpha_{min})f/c_0$$

$$\frac{k_{zmax}}{k_{zmin}} = \frac{1}{\alpha_{min}} = \sqrt{1+R^2}$$

The strategy for choosing frequencies:

$$k_{zmin}(f_{n+1}) = k_{zmax}(f_n)$$

Because of the former $k = {}^{4\pi f\alpha}/_{c_0}$, obtain the relation

$$f_{n+1} = \frac{f_n}{\alpha_{min}}$$

and the frequency increment

$$\Delta f_{n+1} = f_{n+1} - f_n = \left(\frac{1-\alpha_{min}}{\alpha_{min}}\right) f_n = (1-\alpha_{min}) f_{n+1}$$

3.6 The equivalence between gradient images and migration

Migration maps the data to “isochrones” in the model space, whereas the gradient maps the data residuals to the wavepath. Transmitted events map within the first Fresnel zones of the wavepath, while reflected events map to the higher order Fresnel zones.

In the first iteration of a waveform inversion scheme, the starting model is normally a smoothed model, which will generate accurate transmitted arrivals but no reflected energy. The first iteration data residuals will be dominated by reflections, the first iteration image is kinematically equivalent to a migration of the data.

4 Plessix_2010_SEG_Application to land data set[4]

4.1 Introduction

* Proposing of full waveform inversion: Tarantola, 1987.
* 3D real marine examples: Plessix, 2009; Sirgue *et al.*, 2009; Vigh *et al.*, 2009.
* ************************
* Low frequencies and large offsets mitigate the sensitivity to the initial mdoel: Bunks *et al.*, 1995; Pratt, 1999.
* FWI can update the long spatial wavelengths of velocity: Gauthier *et al.*, 1986; Pratt, 1999.
* Apply to land data sets: Ravaut *et al.*, 2004; Brenders and Pratt, 2004. (Attenuate the elastic effects by focusing on the first breaks with windowing technique)
* ************************
* Solve the wave equation in the frequency domain: Plessix, 1997.
* The width of the valleys of the least-squares misfit is inversely proportional to frequency: Bunks *et al.*, 1995.
* Overlap the frequencies between scales to better retain the velocity updates of the low frequency scales: Brossier *et al.*, 2009.

4.2 Full waveform inversion

The misfit function

$$J_f(m) = \frac{1}{2}||W(c-d)||^2$$

with a frequency f, the velocity field m, the modeled data c, the observed data d and a data weighting matrix W which is a diagonal matrix where the diagonal elements are h^β with the offset h and a coefficient β generally between 0 and 2.

Minimize with the quasi-Newton algorithm

$$m_{k+1} = m_k - \alpha_k B_k \nabla_m J_f(m_k)$$

with the step length α_k and the approximated inverse B_k of the Hessian.

5 Fichtner_2010_EPSL_Full waveform tomography[5]

5.1 Introduction

* Simulating of seismic waves with heterogeneous Earth models: Faccioli *et al.*, 1997; Komatitsch and Tromp, 2002; Dumbser and Käser, 2006.
* ************************

[4] René-Edouard Plessix, Guido Baeten and Jan Willem de Maag *et al.*, 2010, SEG 2010 Annual Meeting, Application of acoustic full waveform inversion to a low-frequency large-offset land data set. Date: 2016/10/3 Mon.

[5] Andreas Fichtner, Brian L. N. Kennett and Heiner Igel *et al.*, 2010, Earth and Planetary Science Letters, Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle. Date: 2016/10/13 Thu.

* Full waveform tomography: Konishi *et al.*, 2009; Tape *et al.*, 2009; Fichtner *et al.*, 2009a & 2009b.

* ************************

* Spectral-element method in an Earth model with 3D variations: Fichtner *et al.*, 2009a.

* The discrete equations are solved in parallel: Oeser *et al.*, 2006.

* crust2.0 model : Bassin *et al.*, 2000 (please hit here to download the model data).

* Measure time-frequency phase misfits to extract waveform information: Fichtner *et al.*, 2008.

* The η parameter: Takeuchi and Saito, 1972 (NO Source).

* Set the variations of v_{ph} and v_{pv} to 0.5 times the variations of v_{sh} and v_{sv}: Nettles and Dziewonski, 2008.

* Previous tomography results of the Australasian upper mantle: Debayle and Kennett, 2000a; Fishwick *et al.*, 2005.

* Minimise the cumulative phase misfit using a preconditioned conjugate-gradient method: Fichtner *et al.*, 2009b.

* The adjoint method: Tarantola, 1988; Tromp *et al.*, 2005; Fichtner *et al.*, 2006; Sieminski *et al.*, 2007a & 2007b.

* Refracted body wave studies on Australasian region: Kaiho and Kennett, 2000.

* Elastic 1D reference model PREM: Dziewonski and Anderson, 1981.

* 3D model of shear wave attenuation on Australasian region: Abdulah, 2007.

* Previous surface wave studies on Australia: Zielhuis and van der Hilst, 1996; Simons *et al.*, 1999 & 2002; Debayle and Kennett, 2000a; Yoshizawa and Kennett, 2004; Fishwick *et al.*, 2005.

* Time-frequency phase and amplitude misfits are strongly related: Tian *et al.*, 2009.

* Tomographic study of the radial anisotropy in the Australian region: Debayle and Kennett, 2000a & 2000b.

* Global studies of radial anisotropy: Montagner, 2002; Panning and Romanowicz, 2006; Nettles and Dziewonski, 2008.

* AK135 model : Kennett *et al.*, 1995.

* A Centralian Superbasin existed between 1000 and 750 Ma: Myers *et al.*, 1996.

* SKS splitting studies below Australia: Clitheroe and van der Hilst, 1998.

* Azimuthal anisotropy studies around 150km depth below Australia: Debayle and Kennett, 2000a & 2000b; Simons *et al.*, 2002.

* The Lehmann discontinuity: Lehmann, 1961; Karato, 1992.

* Dislocation creep continues to be dominant to depth of 330km: Mainprince *et al.*, 2005; Raterron *et al.*, 2009.

5.2 Seismic anisotropy

Mineralogical seismic anisotropy (MSA) is the result of the coherent lattice-preferred orientation of anisotropic minerals over length scales that exceed the resolution length. **Structural seismic anisotropy** (SSA) is induced by heterogeneities with length scales that can not be resolved. MSA and SSA can not be distinguished seismologically, but the influence of SSA on the tomographic images can be reduced by increasing the tomographic resolution.

The geodynamic interpretation of seismic anisotropy is based on its relation to flow in the Earth. Horizontal (vertical) flow causes preferentially horizontal (vertical) alignment of small-scale heterogeneities and thus leads to positive (negative) radial SSA, i.e. $v_{sh} > v_{sv}$ ($v_{sh} < v_{sv}$). The development of MSA in the presence of flow depends mostly on the relation between shear strain and the lattice-preferred orientation formation of olivine.

5.3 Filtering of tomographic images

The spatial filtering of regional tomographic images involves: the representation of the images in terms of spherical splines; the application of a spherical convolution.

5.3.1 Spherical spline expansion

A physical quantity m_d is defined at discrete points $\xi_1, \xi_2, \ldots, \xi_N$ that lie within a section Ω_s of the unit sphere Ω. The discretely defined quantity m can be interpolated using a spherical spline of the form

$$m(\mathbf{x}) = \sum_{k=1}^{N} \mu_k K_h(\mathbf{x}, \xi_k), \qquad \mathbf{x}, \xi_1, \xi_2, \ldots, \xi_N \in \Omega_s \subset \Omega$$

where K_h is a spline basis function and when using an Abel-Poisson kernel:

$$K_h(\mathbf{x}, \xi_k) = \frac{1}{4\pi} \frac{1-h^2}{[1+h^2-2h(\mathbf{x}\cdot\xi_k)]^{3/2}}$$

And h is chosen depending on the typical distance between the collocation points ξ_k. μ_k is found through the solution of the linear system of equations:

$$m_d(\xi_i) = m(\xi_i) = \sum_{k=1}^{N} \mu_k K_h(\xi_i, \xi_k), i = 1, 2, \ldots, N$$

5.3.2 Filtering through spherical convolution

Filter a tomographic image by convolving its spherical spline representation, $m(\mathbf{x})$ with a filter function $\phi \in L^2[-1, 1]$:

$$(m * \phi)(\mathbf{x}) = \int_{\Omega} m(\xi)\phi(\xi \cdot \mathbf{x}) d^3\xi$$

The above equation is called the spherical convolution of m with ϕ. When expressed in terms of the Legendre coefficients ϕ_n of ϕ and the spherical harmonic coeffients m_{nj} of $\mathbf{m}$:

$$(m * \phi)(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \phi_n m_{nj} Y_{nj}(\mathbf{x})$$

where Y_{nj} are the spherical harmonic functions of degree n and order j. A filter function ϕ with continuously decreasing Legendre coefficients acts as a low-pass filter.

The Abel-Poisson scaling functions:

$$\phi^{(a)}(t) = \frac{1}{4\pi} \frac{1 - p^2}{(1 + p^2 - 2pt)^{3/2}}, p = e^{-2^{-a}}, a \in N^+$$

where small values of a give low-pass filters and vice versa. The Legendre coefficients $\phi_n^{(a)}$ of $\phi^{(a)}$ are $e^{-n2^{-a}}$. Combining spherical splines and Abel-Poisson scaling functions, obtain:

$$(m * \phi)(\mathbf{x}) = \sum_{k=1}^{N} \mu_k K_{h'}(\mathbf{x}, \xi_k)$$

Thus, the filtering is achieved by simply replacing the parameter h in the original sperical spline with the modified parameter $h' = he^{-2-a}$.

6 Tromp_2005_GJI_Adjoint methods[6]

6.1 Introduction

* Solve iteratively the seismic inverse problem by numerically calculating the Fréchet derivatives of a waveform misfit function & introduce the concept of an adjoint field: Tarantola, 1984 (for the acoustic wave equation) & 1987 & 1988 (for the (an-)elastic wave equation).
* Develop and implement the acoustic theory on seismic inversion: Tarantola, 1984.
* Illustrate numerically the acoustic theory of seismic waveform inversion: Gauthier *et al.*, 1986.
* Extend the acoustic theory to the (an-)elastic wave equation: Tarantola, 1987 & 1988.
* Apply the (an-)elastic theory to real data: Crase *et al.*, 1990.
* Other applications of the solving iteratively theory: Mora, 1987 & 1988; Pratt, 1999; Akcelik *et al.*, 2002 & 2003.
* *************************
* Introduce an 'adjoint' calculation as a means of determining the gradient of a misfit function: Talagrand and Courtier, 1987.
* *************************
* Found the concept of 'time-reversal mirrors' (an acoustic signal is recorded, time-reversed and retransmitted) & time-reversal imaging: Fink *et al.*, 1989; Fink, 1992 & 1997.
* *************************
* Take use of finite-frequency kernels for traveltime or amplitude inversions: Marquering *et al.*, 1999; Zhao *et al.*, 2000; Dahlen *et al.*, 2000; Hung *et al.*, 2000; Dahlen and Baig, 2002.
* Implement finiet-frequency kernels for compressional-wave tomography: Montelli *et al.*, 2004.
* *************************
* The least-squares waveform misfit function: Nolet, 1987.

[6] Jeroen Tromp, Carl Tape and Qinya Liu, 2005, Geophys. J. Int., Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Date: 2016/10/17 Mon.

* Determine Fréchet derivatives based upon the Born approximation: Hudson, 1977; Wu and Aki, 1985.
* A standard conjugate-gradient algorithm: Fletcher and Reeves, 1964; Mora, 1987 & 1988.
* Reconstruct the regular field **s** using the final displacement field $\mathbf{s}(\mathbf{x}, T)$ as a starting point for integration backward in time: Gauthier *et al.*, 1986
* The spectral-element method of seismic wave propagation in anelastic materials: Komatitsch and Tromp, 1999 & 2002a.
* The finite-frequency traveltime tomography: Zhao *et al.*, 2000; Dahlen *et al.*, 2000; Hung *et al.*, 2000.
* The Generalized Seismological Data Functionals (GSDF): Gee and Jordan, 1992 (introduce); Chen *et al.*, 2004 (extend).
* Spectral-element method: Komatitsch and Tromp, 1999.
* The finite-frequency traveltime kernels using ray-based methods: Hung *et al.*, 2000.
* Welch tapering window: Press *et al.*, 1994.

6.2 Waveform tomography

To minimize the differences between waveform data $\mathbf{d}(\mathbf{x}_r, t)$ recorded at N stations $\mathbf{x}_r$, $r = 1, 2, \ldots, N$, and the corresponding synthetics $\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})$ for the current M-dimensional model vector $\mathbf{m}$, introduce the least-squares waveform misfit function:

$$\chi(m) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T ||\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) - \mathbf{d}(\mathbf{x}_r, t)||^2 dt$$

where **d** and **s** can be windowed and filtered on the time interval $[0, T]$. An iterative inversion requires the calculation of the Fréchet derivatives:

$$\delta\chi = \sum_{r=1}^{N}\int_0^T [\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) - \mathbf{d}(\mathbf{x}_r, t)] \cdot \delta\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) dt$$

where $\delta\mathbf{s}$ denotes the perturbation in the displacement field **s** due to a model perturbation $\delta\mathbf{m}$.

In seismic tomography, Fréchet derivatives may be determined based upon the Born approximation. Suppose having a generic background model $\{\rho, c_{jklm}\}$ with perturbations $\{\delta\rho, \delta c_{jklm}\}$, the associated perturbed displacement (the following equation can be referred to eq.2.43 on P.28 of the doctoral thesis of Yan JIANG):

$$\delta s_i(\mathbf{x}, t) = -\int_0^t \int_V [\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}, \mathbf{x}'; t-t')\partial_{t'}^2 s_j(\mathbf{x}', t') + \delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}, \mathbf{x}'; t-t')\partial_l' s_m(\mathbf{x}', t')] d^3\mathbf{x}' dt'$$

where V is the model volume. Obtain:

$$\delta\chi = -\sum_{r=1}^{T}\int_0^T [s_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)] \int_0^t \int_V [\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t')\partial_{t'}^2 s_j(\mathbf{x}', t')$$
$$\delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t')\partial_l' s_m(\mathbf{x}', t')] d^3\mathbf{x}' dt' dt$$

Define the field:

$$\Phi_k(\mathbf{x}', t') = \sum_{r=1}^{N}\int_{t'}^{T} G_{ik}(\mathbf{x}_r, \mathbf{x}'; t-t')[s_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)] dt$$

Using the reciprocity $G_{ik}(\mathbf{x}_r, \mathbf{x}'; t-t') = G_{ki}(\mathbf{x}', \mathbf{x}_r; t-t')$,

$$\Phi_k(\mathbf{x}', t') = \sum_{r=1}^{N} \int_{t'}^{T} G_{ki}(\mathbf{x}', \mathbf{x}_r; t-t')[s_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)]dt$$

Making the substitution $t \to T-t$,

$$\Phi_k(\mathbf{x}', t') = \sum_{r=1}^{N} \int_{0}^{T-t'} G_{ki}(\mathbf{x}', \mathbf{x}_r; T-t-t')[s_i(\mathbf{x}_r, T-t) - d_i(\mathbf{x}_r, T-t)]dt$$

Next define the <u>*waveform adjoint source*</u>:

$$f_i^{\dagger}(\mathbf{x}, t) = \sum_{r=1}^{N} [s_i(\mathbf{x}_r, T-t) - d_i(\mathbf{x}_r, T-t)]\delta(\mathbf{x} - \mathbf{x}_r)$$

With the above definition,

$$\Phi_k(\mathbf{x}', t') = \int_{0}^{T-t'} \int_V G_{ki}(\mathbf{x}', \mathbf{x}; T-t-t') f_i^{\dagger}(\mathbf{x}, t) d^3\mathbf{x}\, dt$$

Take the relationship $\Phi_k(\mathbf{x}', T-t') = s_k^{\dagger}(\mathbf{x}', t')$,

$$s_k^{\dagger}(\mathbf{x}', t') = \int_{0}^{t'} \int_V G_{ki}(\mathbf{x}', \mathbf{x}; t'-t) f_i^{\dagger}(\mathbf{x}, t) d^3\mathbf{x}\, dt$$

where $\mathbf{s}^{\dagger}$ is the introduced <u>*waveform adjoint field*</u> generated by the waveform adjoint source.

With the introduction of the adjoint field,

$$\delta\chi = \int_V [K_{\rho}(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_{c_{jklm}}(\mathbf{x})\delta \ln c_{jklm}(\mathbf{x})]d^3\mathbf{x}$$

where $\delta \ln \rho = {}^{\delta\rho}/_{\rho}$ and $\delta \ln c_{jklm} = {}^{\delta c_{jklm}}/_{c_{jklm}}$ denote relative model perturbations, and the 3-D <u>*waveform misfit kernels*</u> for density and the elastic parameters are respectively:

$$K_{\rho}(\mathbf{x}) = -\int_0^T \rho(\mathbf{x})\mathbf{s}^{\dagger}(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t)dt$$

$$K_{c_{jklm}}(\mathbf{x}) = -\int_0^T \epsilon_{jk}^{\dagger}(\mathbf{x}, T-t)c_{jklm}(\mathbf{x})\epsilon_{lm}(\mathbf{x}, t)dt$$

where $\epsilon^{\dagger} = {}^{1}/_{2}[\nabla \mathbf{s}^{\dagger} + (\nabla \mathbf{s}^{\dagger})^T]$, ϵ_{lm} and $\epsilon_{jk}^{\dagger}$ denote the strain and the waveform adjoint strain tensors, respectively.

For an isotropic matreical, $c_{jklm} = (\kappa - {}^{2\mu}/_{3})\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})$, thus

$$\delta\chi = \int_V [K_{\rho}(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_{\mu}(\mathbf{x})\delta \ln \mu(\mathbf{x}) + K_{\kappa}(\mathbf{x})\delta \ln \kappa(\mathbf{x})]d^3\mathbf{x}$$

where the <u>*isotropic misfit kernels*</u> K_{μ} and K_{κ} for the bulk and shear moduli κ and μ are respectively:

$$K_{\mu}(\mathbf{x}) = -\int_0^T 2\mu(\mathbf{x})\mathbf{D}^{\dagger}(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t)dt$$

$$K_{\kappa}(\mathbf{x}) = -\int_0^T \kappa(\mathbf{x})[\nabla \cdot \mathbf{s}^{\dagger}(\mathbf{x}, T-t)][\nabla \cdot \mathbf{s}(\mathbf{x}, t)]dt$$

where $\mathbf{D}$ and $\mathbf{D}^{\dagger}$ denote the traceless strain deviator and its waveform adjoint, respectively.

Alternatively,

$$\delta\chi = \int_V [K'_{\rho}(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_{\beta}(\mathbf{x})\delta \ln \beta(\mathbf{x}) + K_{\alpha}(\mathbf{x})\delta \ln \alpha(\mathbf{x})]d^3\mathbf{x}$$

$$K'_{\rho} = K_{\rho} + K_{\kappa} + K_{\mu}, \qquad K_{\beta} = 2\Big(K_{\mu} - \frac{4\mu}{3\kappa}K_{\kappa}\Big), \qquad K_{\alpha} = 2\Big(\frac{\kappa + (4/3)\mu}{\kappa}\Big)K_{\kappa}$$

6.2.1 Topography on internal discontinuities

Let δh denote topographic perturbations in the direction of the unitoutward normal $\hat{\mathbf{n}}$ on solid-solid discontinuities Σ_{SS} or fluid-solid discontinuities Σ_{FS}, the perturbed displacement field $\delta\mathbf{s}$ due to topographic perturbations δh (Dahlen, 2004):

$$\begin{aligned}\delta s_i(\mathbf{x},t) = \int_0^t \int_\Sigma &[\rho(\mathbf{x}')G_{ij}(\mathbf{x},\mathbf{x}';t-t')\partial_{t'}^2 s_j(\mathbf{x}',t') + \partial_k' G_{ij}(\mathbf{x},\mathbf{x}';t-t')c_{jklm}(\mathbf{x}')\partial_l' s_m(\mathbf{x}',t')\\ &- \hat{n}_k(\mathbf{x}')\partial_n' G_{ij}(\mathbf{x},\mathbf{x}';t-t')c_{jklm}(\mathbf{x}')\partial_l' s_m(\mathbf{x}',t')\\ &- \hat{n}_k(\mathbf{x}')c_{jklm}(\mathbf{x}')\partial_l' G_{im}(\mathbf{x},\mathbf{x}';t-t')\partial_n' s_j(\mathbf{x}',t')]_-^+ \delta h(\mathbf{x}')d^2\mathbf{x}'dt'\\ &+ \int_0^t \int_{\Sigma_{FS}} [G_{ik}(\mathbf{x},\mathbf{x}';t-t')\hat{n}_j(\mathbf{x}')\hat{n}_p(\mathbf{x}')c_{jplm}(\mathbf{x}')\partial_l' s_m(\mathbf{x}',t')\\ &+ s_k(\mathbf{x}',t')\hat{n}_j(\mathbf{x}')\hat{n}_p(\mathbf{x}')c_{jplm}(\mathbf{x}')\partial_l' G_{im}(\mathbf{x},\mathbf{x}';t-t')]_-^+ \nabla_k^{\Sigma'} \delta h(\mathbf{x}')d^2\mathbf{x}'dt'\end{aligned}$$

where $\Sigma = \Sigma_{SS} + \Sigma_{FS}$ denote all discontinuities, the surface gradient $\nabla^\Sigma = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \nabla$ and the normal derivative $\partial_n = \hat{\mathbf{n}} \cdot \nabla$. Therefore, the gradient of the misfit function due to topographic perturbations δh:

$$\delta\chi = \int_\Sigma K_h(\mathbf{x})\delta h(\mathbf{x})d^2\mathbf{x} + \int_{\Sigma_{FS}} \mathbf{K}_h(\mathbf{x}) \cdot \nabla^\Sigma \delta h(\mathbf{x})d^2\mathbf{x}$$

where

$$\begin{aligned}K_h(\mathbf{x}) = \int_0^T &[\rho(\mathbf{x})\mathbf{s}^\dagger(\mathbf{x},T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x},t) + \epsilon^\dagger(\mathbf{x},T-t) : \mathbf{c}(\mathbf{x}) : \epsilon(\mathbf{x},t)\\ &- \hat{\mathbf{n}}(\mathbf{x})\partial_n \mathbf{s}^\dagger(\mathbf{x},T-t) : \mathbf{c}(\mathbf{x}) : \epsilon(\mathbf{x},t) - \hat{\mathbf{n}}(\mathbf{x})\partial_n \mathbf{s}(\mathbf{x},t) : \mathbf{c}(\mathbf{x}) : \epsilon^\dagger(\mathbf{x},T-t)]_-^+ dt\end{aligned}$$

$$\mathbf{K}_h(\mathbf{x}) = \int_0^T [\mathbf{s}^\dagger(\mathbf{x},T-t)\hat{\mathbf{n}}(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) : \mathbf{c}(\mathbf{x}) : \epsilon(\mathbf{x},t) + \mathbf{s}(\mathbf{x},t)\hat{\mathbf{n}}(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) : \mathbf{c}(\mathbf{x}) : \epsilon^\dagger(\mathbf{x},T-t)]_-^+ dt$$

In an isotropic earth model,

$$\begin{aligned}K_h(\mathbf{x}) = \int_0^T &[\rho(\mathbf{x})\mathbf{s}^\dagger(\mathbf{x},T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x},t) + \kappa(\mathbf{x})\nabla \cdot \mathbf{s}^\dagger(\mathbf{x},T-t)\nabla \cdot \mathbf{s}(\mathbf{x},t) + 2\mu(\mathbf{x})\mathbf{D}^\dagger(\mathbf{x},T-t) : \mathbf{D}(\mathbf{x},t)\\ &- \kappa(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) \cdot \partial_n \mathbf{s}^\dagger(\mathbf{x},T-t)\nabla \cdot \mathbf{s}(\mathbf{x},t) - 2\mu(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x})\partial_n \mathbf{s}^\dagger(\mathbf{x},T-t) : \mathbf{D}(\mathbf{x},t)\\ &- \kappa(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) \cdot \partial_n \mathbf{s}(\mathbf{x},t)\nabla \cdot \mathbf{s}^\dagger(\mathbf{x},T-t) - 2\mu(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x})\partial_n \mathbf{s}(\mathbf{x},t) : \mathbf{D}^\dagger(\mathbf{x},T-t)]_-^+ dt\end{aligned}$$

$$\begin{aligned}\mathbf{K}_h(\mathbf{x}) = \int_0^T &[\mathbf{s}^\dagger(\mathbf{x},T-t)[\kappa(\mathbf{x})\nabla \cdot \mathbf{s}(\mathbf{x},t) + 2\mu(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x},t) \cdot \hat{\mathbf{n}}(\mathbf{x})]\\ &+ \mathbf{s}(\mathbf{x},t)[\kappa(\mathbf{x})\nabla \cdot \mathbf{s}^\dagger(\mathbf{x},T-t) + 2\mu(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) \cdot \mathbf{D}^\dagger(\mathbf{x},T-t) \cdot \hat{\mathbf{n}}(\mathbf{x})]]_-^+ dt\end{aligned}$$

Besides, on the Earth's free surface the traction vanishes,

$$K_h(\mathbf{x}) = -\int_0^T [\rho(\mathbf{x})\mathbf{s}^\dagger(\mathbf{x},T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x},t) + \epsilon^\dagger(\mathbf{x},T-t) : \mathbf{c}(\mathbf{x}) : \epsilon(\mathbf{x},t)]dt$$

In the isotropic case,

$$K_h(\mathbf{x}) = -\int_0^T [\rho(\mathbf{x})\mathbf{s}^\dagger(\mathbf{x},T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x},t) + \kappa(\mathbf{x})\nabla \cdot \mathbf{s}(\mathbf{x},t)\nabla \cdot \mathbf{s}^\dagger(\mathbf{x},T-t) + 2\mu(\mathbf{x})\mathbf{D}(\mathbf{x},t) : \mathbf{D}^\dagger(\mathbf{x},T-t)]dt$$

6.3 Adjoint equations

The equation of motion in an anelastic earth model:

$$\rho\partial_t^2\mathbf{s} = \nabla\cdot\mathbf{T} + \mathbf{f}$$

where $\mathbf{T}$ is the symmetric stress tensor in an anelastic matreical. For the unrelaxed elastic tensor $\mathbf{c}^U$, the displacement gradient $\nabla\mathbf{s}$ and L symmetric memory variable tensors $\mathbf{R}^l$, $l = 1, 2, \ldots, L$,

$$\mathbf{T} = \mathbf{c}^U : \nabla\mathbf{s} - \sum_{l=1}^{L}\mathbf{R}^l$$

where $\mathbf{R}^l$ represent standard linear solids. For each standard linear solid,

$$\partial_t\mathbf{R}^l = -\frac{\mathbf{R}^l}{\tau^{\sigma l}} + \frac{\delta\mathbf{c}^l : \nabla\mathbf{s}}{\tau^{\sigma l}}$$

The above equations need to be subject to the boundary conditions, that on the stress-free surface $\hat{\mathbf{n}}\cdot\mathbf{T} = 0$ and at solid-solid boundaries both $\mathbf{s}$ and $\hat{\mathbf{n}}\cdot\mathbf{T}$ are continuous whereas at fluid-solid boundaries both $\hat{\mathbf{n}}\cdot\mathbf{s}$ and $\hat{\mathbf{n}}\cdot\mathbf{T}$ are continuous. In terms of the relaxed modulus c^R_{ijkl},

$$c^U_{ijkl} = c^R_{ijkl}\Big(1 - \sum_{l=1}^{L}\big(1 - \frac{\tau^{\epsilon l}_{ijkl}}{\tau^{\sigma l}}\big)\Big)$$

where $\tau^{\epsilon l}_{ijkl}$ and $\tau^{\sigma l}$ are the strain and stress relaxation times, respectively. The modulus defect $\delta\mathbf{c}^l$:

$$\delta c^l_{ijkl} = -c^R_{ijkl}\big(1 - \frac{\tau^{\epsilon l}_{ijkl}}{\tau_{\sigma l}}\big)$$

Replace the source $\mathbf{f}$ with the waveform adjoint source $\mathbf{f}^\dagger$, obtain the adjoint equations:

$$\rho\partial_t^2\mathbf{s}^\dagger = \nabla\cdot\mathbf{T}^\dagger + \mathbf{f}^\dagger$$

$$\mathbf{T}^\dagger = \mathbf{c}^U : \nabla\mathbf{s}^\dagger - \sum_{l=1}^{L}\mathbf{R}^{l\dagger}$$

$$\partial_t\mathbf{R}^{l\dagger} = -\frac{\mathbf{R}^{l\dagger}}{\tau^{\sigma l}} + \frac{\delta\mathbf{c}^l : \nabla\mathbf{s}^\dagger}{\tau^{\sigma l}}$$

For completeness, the adjoint momentum equation for a rotating, self-gravitating Earth model is:

$$\rho(\partial_t^2\mathbf{s}^\dagger - 2\Omega\times\partial_t\mathbf{s}^\dagger) = \nabla\cdot\mathbf{T}^\dagger + \nabla(\rho\mathbf{s}^\dagger\cdot\mathbf{g}) - \rho\nabla\phi^\dagger - \nabla\cdot(\rho\mathbf{s}^\dagger)\mathbf{g} + \mathbf{f}^\dagger$$

where Ω and $\mathbf{g}$ denote the angular velocity and the equilibrium gravitational acceleration of the earth model, respectively.

6.4 Traveltime tomography

Introduce the traveltime misfit function

$$\chi(m) = \frac{1}{2}\sum_{r=1}^{N}[T_r(m) - T_r^{obs}]^2$$

where $T_r(m)$ and T_r^{obs} denote the predicted and observed traveltime at station r, respectively. The gradient of the misfit function is:

$$\delta\chi = \sum_{r=1}^{N}[T_r(m) - T_r^{obs}]\delta T_r$$

6.4.1 Banana-doughnut kernels

The Fréchet derivative of the traveltime in terms of the cross-correlation of an observed and synthetic waveform (refer to eq.2.46 on P.31 of the doctoral thesis of Yan JIANG):

$$\delta T_r = \frac{1}{N_r}\int_0^T w_r(t)\partial_t s_i(\mathbf{x}_r,t)\delta s_i(\mathbf{x}_r,t)dt$$

$$N_r = \int_0^T w_r(t)s_i(\mathbf{x}_r,t)\partial_t^2 s_i(\mathbf{x}_r,t)dt$$

where w_r denotes the cross-correlation window and δs_i the displacement perturbation. After substitution of $\delta\mathbf{s}$ based on the Born approximation,

$$\begin{aligned}\delta T_r = &-\frac{1}{N_r}\int_0^T w_r(t)\partial_t s_i(\mathbf{x}_r,t)\int_0^t\int_V[\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}_r,\mathbf{x}';t-t')\partial_{t'}^2 s_j(\mathbf{x}',t')\\ &+\delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}_r,\mathbf{x}';t-t')\partial_l' s_m(\mathbf{x}',t')]d^3\mathbf{x}'dt'dt\end{aligned}$$

Define the _traveltime adjoint source_ $\bar{\mathbf{f}}^\dagger$ and the _traveltime adjoint field_ $\bar{\mathbf{s}}^\dagger$:

$$\bar{f}_i^\dagger(\mathbf{x},t) = \frac{1}{N_r}w_r(T-t)\partial_t s_i(\mathbf{x}_r,T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

$$\begin{aligned}\bar{s}_j^\dagger(\mathbf{x}',\mathbf{x}_r,T-t') &= \int_0^{T-t'} G_{ji}(\mathbf{x}',\mathbf{x};T-t-t')\bar{f}_i^\dagger(\mathbf{x},t)dt\\ &= \frac{1}{N_r}\int_0^{T-t'} G_{ji}(\mathbf{x}',\mathbf{x}_r;T-t-t')w_r(T-t)\partial_t s_i(\mathbf{x}_r,T-t)dt\end{aligned}$$

With this definition the isotropic traveltime Fréchet derivative is:

$$\delta T_r = \int_V[\bar{K}_\rho(\mathbf{x},\mathbf{x}_r)\delta\ln\rho(\mathbf{x}) + \bar{K}_\mu(\mathbf{x},\mathbf{x}_r)\delta\ln\mu(\mathbf{x}) + \bar{K}_\kappa(\mathbf{x},\mathbf{x}_r)\delta\ln\kappa(\mathbf{x})]d^3\mathbf{x}$$

where the _banana-doughnut kernels_ $\bar{K}_\rho$, $\bar{K}_\mu$ and $\bar{K}_\kappa$ are:

$$\bar{K}_\rho(\mathbf{x},\mathbf{x}_r) = -\int_0^T \rho(\mathbf{x})[\bar{\mathbf{s}}^\dagger(\mathbf{x},\mathbf{x}_r,T-t)\cdot\partial_t^2\mathbf{s}(\mathbf{x},t)]dt$$

$$\bar{K}_\mu(\mathbf{x},\mathbf{x}_r) = -\int_0^T 2\mu(\mathbf{x})\bar{\mathbf{D}}^\dagger(\mathbf{x},\mathbf{x}_r,T-t):\mathbf{D}(\mathbf{x},t)dt$$

$$\bar{K}_\kappa(\mathbf{x},\mathbf{x}_r) = -\int_0^T \kappa(\mathbf{x})[\nabla\cdot\bar{\mathbf{s}}^\dagger(\mathbf{x},\mathbf{x}_r,T-t)][\nabla\cdot\mathbf{s}(\mathbf{x},t)]dt$$

where $\bar{\mathbf{D}}^\dagger$ denotes the traveltime adjoint strain deviator associated with $\bar{\mathbf{s}}^\dagger$. Alternatively,

$$\delta T_r = \int_V[\bar{K}_\rho'(\mathbf{x},\mathbf{x}_r)\delta\ln\rho(\mathbf{x}) + \bar{K}_\beta(\mathbf{x},\mathbf{x}_r)\delta\ln\beta(\mathbf{x}) + \bar{K}_\alpha(\mathbf{x},\mathbf{x}_r)\delta\ln\alpha(\mathbf{x})]d^3\mathbf{x}$$

$$\bar{K}_\rho' = \bar{K}_\rho + \bar{K}_\kappa + \bar{K}_\mu, \qquad \bar{K}_\beta = 2\Big(\bar{K}_\mu - \frac{4\mu}{3\kappa}\bar{K}_\kappa\Big), \qquad \bar{K}_\alpha = 2\Big(1+\frac{4\mu}{3\kappa}\Big)\bar{K}_\kappa$$

6.4.2 Misfit kernels

The Fréchet derivative of the traveltime misfit function is:

$$\begin{aligned}\delta\chi &= \sum_{r=1}^{N}(T_r - T_r^{obs})\delta T_r \\ &= \int_V [K'_\rho(\mathbf{x})\delta\ln\rho(\mathbf{x}) + K_\beta(\mathbf{x})\delta\ln\beta(\mathbf{x}) + K_\alpha(\mathbf{x})\delta\ln\alpha(\mathbf{x})]d^3\mathbf{x}\end{aligned}$$

where the *traveltime misfit kernels* K'_ρ, K_β and K_α are:

$$K'_\rho(\mathbf{x}) = \sum_{r=1}^{N}(T_r - T_r^{obs})\bar{K}'_\rho(\mathbf{x}, \mathbf{x}_r)$$

$$K_\beta(\mathbf{x}) = \sum_{r=1}^{N}(T_r - T_r^{obs})\bar{K}_\beta(\mathbf{x}, \mathbf{x}_r)$$

$$K_\alpha(\mathbf{x}) = \sum_{r=1}^{N}(T_r - T_r^{obs})\bar{K}_\alpha(\mathbf{x}, \mathbf{x}_r)$$

Define the *combined traveltime adjoint field* $\mathbf{s}^\dagger$ and the *combined traveltime adjoint source* $\mathbf{f}^\dagger$:

$$\mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N}(T_r - T_r^{obs})\bar{\mathbf{s}}^\dagger(\mathbf{x}, \mathbf{x}_r, t)$$

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N}(T_r - T_r^{obs})\frac{1}{N_r}w_r(T-t)\partial_t s_i(\mathbf{x}_r, T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

6.4.3 Differential traveltime tomography

Suppose observed differential traveltimes ΔT_r^{obs} and predicted differential traveltimes $\Delta T_r(m) = T_r^A(m) - T_r^B(m)$ between two phases A and B for the station r ($r = 1, 2, \ldots, N$). Minimize the *differential traveltime misfit function* and its gradient are:

$$\chi(m) = \frac{1}{2}\sum_{r=1}^{N}[\Delta T_r(m) - \Delta T_r^{obs}]^2$$

$$\delta\chi = \sum_{r=1}^{N}[\Delta T_r(m) - \Delta T_r^{obs}]\delta\Delta T_r$$

where $\delta\Delta T_r = \delta T_r^A - \delta T_r^B$.

Define the *combined differential traveltime adjoint field* $\Delta\mathbf{s}^\dagger$ and the *combined differential traveltime adjoint source* $\mathbf{f}^\dagger$:

$$\Delta\mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N}(\Delta T_r - \Delta T_r^{obs})[\bar{\mathbf{s}}^{A\dagger}(\mathbf{x}, \mathbf{x}_r, t) - \bar{\mathbf{s}}^{B\dagger}(\mathbf{x}, \mathbf{x}_r, t)]$$

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N}(\Delta T_r - \Delta T_r^{obs})\Big[\frac{1}{N_r^A}w_r^A(T-t)\partial_t s_i^A(\mathbf{x}_r, T-t) - \frac{1}{N_r^B}w_r^B(T-t)\partial_t s_i^B(\mathbf{x}_r, T-t)\Big]\delta(\mathbf{x}-\mathbf{x}_r)$$

6.5 Amplitude tomography

Let A_r^{obs} and $A_r(m)$ denote the observed and predicted amplitude of a particular body-wave arrival at the station r, introduce the *amplitude misfit function*:

$$\chi(m) = \frac{1}{2}\sum_{r=1}^{N}\Big[\frac{A_r^{obs}}{A_r(m)} - 1\Big]^2$$

and its gradient is:

$$\delta\chi = \sum_{r=1}^{N}\Big[\frac{A_r^{obs}}{A_r(m)} - 1\Big]\delta\ln A_r$$

The amplitude Fréchet derivative is (Dahlen and Baig, 2002):

$$\delta\ln A_r = \frac{1}{M_r}\int_0^T w_r(t)s_i(\mathbf{x}_r,t)\delta s_i(\mathbf{x}_r,t)dt$$

$$M_r = \int_0^T w_r(t)s_i^2(\mathbf{x}_r,t)dt$$

where w_r denotes the cross-correlation window and δs_i the displacement perturbation.

Define the *amplitude adjoint source* $\bar{\mathbf{f}}^\dagger$ and the *amplitude adjoint field* $\bar{\mathbf{s}}^\dagger$:

$$\bar{f}_i^\dagger(\mathbf{x},t) = \frac{1}{M_r}w_r(T-t)s_i(\mathbf{x}_r,T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

$$\begin{aligned}\bar{s}_j^\dagger(\mathbf{x}',\mathbf{x}_r,T-t') &= \int_0^{T-t'} G_{ji}(\mathbf{x}',\mathbf{x};T-t-t')\bar{f}_i^\dagger(\mathbf{x},t)dt \\ &= \frac{1}{M_r}\int_0^{T-t'} G_{ji}(\mathbf{x}',\mathbf{x}_r;T-t-t')w_r(T-t)s_i(\mathbf{x}_r,T-t)dt\end{aligned}$$

And in terms of the *amplitude kernels* $\bar{K}'_\rho$, $\bar{K}_\beta$ and $\bar{K}_\alpha$,

$$\delta\ln A_r = \int_V[\bar{K}'_\rho(\mathbf{x},\mathbf{x}_r)\delta\ln\rho'(\mathbf{x}) + \bar{K}_\beta(\mathbf{x},\mathbf{x}_r)\delta\ln\beta(\mathbf{x}) + \bar{K}_\alpha(\mathbf{x},\mathbf{x}_r)\delta\ln\alpha(\mathbf{x})]d^3\mathbf{x}$$

Define the *combined amplitude adjoint field* $\mathbf{s}^\dagger$ and the *combined amplitude adjoint source* $\mathbf{f}^\dagger$:

$$\mathbf{s}^\dagger(\mathbf{x},t) = \sum_{r=1}^{N}\Big(\frac{A_r^{obs}}{A_r} - 1\Big)\bar{\mathbf{s}}^\dagger(\mathbf{x},\mathbf{x}_r,t)$$

$$f_i^\dagger(\mathbf{x},t) = \sum_{r=1}^{N}\Big(\frac{A_r^{obs}}{A_r} - 1\Big)\frac{1}{M_r}w_r(T-t)s_i(\mathbf{x}_r,T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

And

$$\delta\chi = \int_V[K'_\rho(\mathbf{x})\delta\ln\rho(\mathbf{x}) + K_\beta(\mathbf{x})\delta\ln\beta(\mathbf{x}) + K_\alpha(\mathbf{x})\delta\ln\alpha(\mathbf{x})]d^3\mathbf{x}$$

6.5.1 Attenuation

For an absorption-band solid, the shear modulus μ is (Liu *et al.*, 1976):

$$\mu(\omega) = \mu(\omega_0)\Big[1 + \frac{2}{\pi}Q_\mu^{-1}\ln\frac{|\omega|}{\omega_0} - i\mathrm{sgn}(\omega)Q_\mu^{-1}\Big]$$

where ω_0 denotes the reference angular frequency. The change in the shear modulus $\delta\mu$ due to perturbations in shear attenuation δQ_μ^{-1} is:

$$\delta\mu(\omega) = \mu(\omega_0)\Big[\frac{2}{\pi}\ln\frac{|\omega|}{\omega_0} - i\mathrm{sgn}(\omega)\Big]\delta Q_\mu^{-1}$$

Taking use of the Born approximation, define the wavefield:

$$\psi_i(\mathbf{x}, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[\frac{2}{\pi}\ln\frac{|\omega|}{\omega_0} - i\mathrm{sgn}(\omega)\Big]s_i(\mathbf{x}, \omega)e^{i\omega t}d\omega$$

and introduce the *Q adjoint source*:

$$\bar{f}_i^\dagger(\mathbf{x}, t) = \frac{1}{M_r}w_r(T - t\psi_i(\mathbf{x}_r, T - t)\delta(\mathbf{x} - \mathbf{x}_r)$$

thus the amplitude anomaly is:

$$\delta\ln A_r = \int_V \bar{K}_\mu(\mathbf{x}, \mathbf{x}_r)\delta Q_\mu^{-1}(\mathbf{x})d^3\mathbf{x}$$

Introduce the *combined Q adjoint source*:

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N}\Big(\frac{A_r^{obs}}{A_r} - 1\Big)\frac{1}{M_r}w_r(T - t)\psi_i(\mathbf{x}_r, T - t)\delta(\mathbf{x} - \mathbf{x}_r)$$

thus the gradient of the attenuation misfit function is:

$$\delta\chi = \int_V K_\mu(\mathbf{x})\delta Q_\mu^{-1}(\mathbf{x})d^3\mathbf{x}$$

6.6 Generalizations

Let $\tau_r(\omega_\lambda)$ denote the frequency-dependent either traveltime anomaly τ_p or amplitude anomaly τ_q at receiver r ($r = 1, 2, \ldots, N$) determined at L discrete angular frequencies ω_λ ($\lambda = 1, 2, \ldots, L$) for the current model m, define the *GSDF misfit function*:

$$\chi(m) = \frac{1}{2}\sum_{r=1}^{N}\sum_{\lambda=1}^{L}[\tau_r(\omega_\lambda)]^2$$

and its gradient is:

$$\delta\chi = \sum_{r=1}^{N}\sum_{\lambda=1}^{L}\tau_r(\omega_\lambda)\delta\tau_r(\omega_\lambda)$$

6.6.1 Banana-doughnut kernels

The time-dependent function $\Psi_i(\mathbf{x}_r, t, \omega_\lambda)$ relates the GSDF parameter perturbations $\delta\tau_r(\omega_\lambda)$ to the seismogram perturbations δs_i:

$$\delta\tau_r(\omega_\lambda) \int_0^T \Psi_i(\mathbf{x}_r, t, \omega_\lambda)\delta s_i(\mathbf{x}_r, t)dt$$

After substitution of $\delta\mathbf{s}$ based on the Born approximation,

$$\begin{aligned}\delta\tau_r(\omega_\lambda) = -\int_0^T \Psi_i(\mathbf{x}_r, t, \omega_\lambda) \int_0^t \int_V [\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t')\partial_{t'}^2 s_j(\mathbf{x}', t') \\ + \delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t')\partial_l' s_m(\mathbf{x}', t')]d^3\mathbf{x}'dt'dt\end{aligned}$$

Define the _GSDF adjoint source_ $\bar{\mathbf{f}}^\dagger$ and the _GSDF adjoint field_ $\bar{\mathbf{s}}^\dagger$:

$$\bar{f}_i^\dagger(\mathbf{x}, t) = \Psi_i(\mathbf{x}_r, T-t, \omega_\lambda)\delta(\mathbf{x}-\mathbf{x}_r)$$

$$\begin{aligned}\bar{s}_j^\dagger(\mathbf{x}', \mathbf{x}_r, T-t', \omega_\lambda) &= \int_0^{T-t'} G_{ji}(\mathbf{x}', \mathbf{x}; T-t-t')\bar{f}_i^\dagger(\mathbf{x}, t)dt \\ &= \int_0^{T-t'} G_{ji}(\mathbf{x}', \mathbf{x}_r; T-t-t')\Psi_i(\mathbf{x}_r, T-t, \omega_\lambda)dt\end{aligned}$$

thus

$$\delta\tau_r(\omega_\lambda) = \int_V [\bar{K}_\rho(\mathbf{x}, \mathbf{x}_r, \omega_\lambda)\delta\ln\rho(\mathbf{x}) + \bar{K}_\mu(\mathbf{x}, \mathbf{x}_r, \omega_\lambda)\delta\ln\mu(\mathbf{x}) + \bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r, \omega_\lambda)\delta\ln\kappa(\mathbf{x})]d^3\mathbf{x}$$

where the _GSDF kernels_ $\bar{K}_\rho$, $\bar{K}_\mu$ and $\bar{K}_\kappa$ are:

$$\bar{K}_\rho(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) = -\int_0^T \rho(\mathbf{x})[\bar{\mathbf{s}}^\dagger(\mathbf{x}, \mathbf{x}_r, T-t, \omega_\lambda)\cdot\partial_t^2\mathbf{s}(\mathbf{x}, t)]dt$$

$$\bar{K}_\mu(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) = -\int_0^T 2\mu(\mathbf{x})\bar{\mathbf{D}}^\dagger(\mathbf{x}, \mathbf{x}_r, T-t, \omega_\lambda) : \mathbf{D}(\mathbf{x}, t)dt$$

$$\bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) = -\int_0^T \kappa(\mathbf{x})[\nabla\cdot\bar{\mathbf{s}}^\dagger(\mathbf{x}, \mathbf{x}_r, T-t, \omega_\lambda)][\nabla\cdot\mathbf{s}(\mathbf{x}, t)]dt$$

where $\bar{\mathbf{D}}^\dagger$ denotes the GSDF adjoint strain deviator associated with the GSDF adjoint field.

6.6.2 Misfit kernels

Introduce the _combined GSDF adjoint field_ $\mathbf{s}^\dagger$ and the _combined GSDF adjoint source_ $\mathbf{f}^\dagger$:

$$\mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda)\bar{\mathbf{s}}^\dagger(\mathbf{x}, \mathbf{x}_r, t, \omega_\lambda)$$

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda)\Psi_i(\mathbf{x}_r, T-t, \omega_\lambda)\delta(\mathbf{x}-\mathbf{x}_r)$$

and the gradient is:

$$\delta\chi = \int_V [K_\rho(\mathbf{x})\delta\ln\rho(\mathbf{x}) + K_\mu(\mathbf{x})\delta\ln\mu(\mathbf{x}) + K_\kappa(\mathbf{x})\delta\ln\kappa(\mathbf{x})]d^3\mathbf{x}$$

where the <u>*combined GSDF kernels*</u> K_ρ, K_μ and K_κ are:

$$K_\rho(\mathbf{x}) = \sum_{r=1}^{N}\sum_{\lambda=1}^{L} \tau_r(\omega_\lambda)\bar{K}_\rho(\mathbf{x},\mathbf{x}_r,\omega_\lambda)$$

$$K_\mu(\mathbf{x}) = \sum_{r=1}^{N}\sum_{\lambda=1}^{L} \tau_r(\omega_\lambda)\bar{K}_\mu(\mathbf{x},\mathbf{x}_r,\omega_\lambda)$$

$$K_\kappa(\mathbf{x}) = \sum_{r=1}^{N}\sum_{\lambda=1}^{L} \tau_r(\omega_\lambda)\bar{K}_\kappa(\mathbf{x},\mathbf{x}_r,\omega_\lambda)$$

6.7 Source inversions

The response $\mathbf{s}(\mathbf{x}, t)$ due to a finite source represented by a moment-density distribution $\mathbf{m}(\mathbf{x}, t)$ on a fault plane Σ is (Dahlen and Tromp, 1998):

$$s_i(\mathbf{x}, t) = \int_0^t \int_\Sigma \partial_j' G_{ik}(\mathbf{x}, \mathbf{x}'; t - t') m_{jk}(\mathbf{x}', t') d^2\mathbf{x}' dt'$$

thus the change $\delta\mathbf{s}$ due to the perturbation $\delta\mathbf{m}$ is:

$$\delta s_i(\mathbf{x}, t) = \int_0^t \int_\Sigma \partial_j' G_{ik}(\mathbf{x}, \mathbf{x}'; t - t') \delta m_{jk}(\mathbf{x}', t') d^2\mathbf{x}' dt'$$

So the previous Fréchet derivative of the waveform misfit function is recast into:

$$\delta\chi = \int_0^T \int_\Sigma \epsilon^\dagger(\mathbf{x}, T - t) : \delta\mathbf{m}(\mathbf{x}, t) d^2\mathbf{x} dt$$

For a point source located at $\mathbf{x}_s$ with the centroid-moment tensor $\mathbf{M}(t)$,

$$\delta\chi = \int_0^T \epsilon^\dagger(\mathbf{x}_s, T - t) : \delta\mathbf{M}(\mathbf{x}_s, t) dt$$

6.8 Joint inversions

For example, the waveform misfit function may be jointly minimized with structural, topographic and source parameters. In that case, its gradient involves perturbations $\delta\mathbf{s}$ due to structural, topographic and source parameters:

$$\begin{aligned}\delta\chi = &\int_V [K_\rho'(\mathbf{x})\delta\ln\rho(\mathbf{x}) + K_\beta(\mathbf{x})\delta\ln\beta(\mathbf{x}) + K_\alpha(\mathbf{x})\delta\ln\alpha(\mathbf{x})] d^3\mathbf{x} \\ &+ \int_\Sigma K_h(\mathbf{x})\delta h(\mathbf{x}) d^2\mathbf{x} + \int_{\Sigma_{FS}} \mathbf{K}_h(\mathbf{x}) \cdot \nabla^\Sigma \delta h(\mathbf{x}) d^2\mathbf{x} \\ &+ \int_0^T \int_\Sigma \epsilon^\dagger(\mathbf{x}, T - t) : \delta\mathbf{m}(\mathbf{x}, t) d^2\mathbf{x} dt\end{aligned}$$

7 Tape_2009_S_Adjoint at SCC[7]

7.1 Introduction

* Seismic tomography adopts PREM model to produce images of Earth's interior: [1] Woodhouse and Dziewonski, 1984 & [2] Romanowicz, 2003 in the mantle; [3] Grand *et al.*, 1997 in subducting slabs; [4] Montelli *et al.*, 2004 in mantle plumes.

* Start the minimization procedure with more realistic 3D initial models: [6] Komatitsch *et al.*, 2002; [7] Akcelik *et al.*, 2003; [8] Chen *et al.*, 2007; [9] Fichtner *et al.*, 2008.

* Use adjoint method within the minimization problem: [10] Tarantola, 1984; [11] Talagrand *et al.*, 1987; [12] Tromp *et al.*, 2005.

* Adjoint tomography: [13] Tape *et al.*, 2007.

* *************************

* 3D seismological model of the southern California crust: [14] Süss and Shaw, 2003; [15] Komatitsch *et al.*, 2004.

* SEM wave propagation code: [15] Komatitsch *et al.*, 2004; [17] Liu and Tromp, 2006 (modify to facilitate an inverse problem).

* *************************

* Empirical relations between elastic wavespeeds and density in Crust: [20] Brocher, 2005.

* *************************

* FLEXWIN windowing code - automated time-window selection algorithm for seismic tomography: [21] Maggi *et al.*, 2009 (please hit here to download the package).

* *************************

* Multi-taper method to make travel-time measurement: [S8] Zhou *et al.*, 2004.

* Subspace methods for multiple parameter classes: [S11] Kennett *et al.*, 1988; [S12] Sambridge *et al.*, 1991.

* Locate southern California seismicity from 1981 to 2005: [S16] Lin *et al.*, 2007.

* Double-difference earthquake location algorithm: [S20] Waldhauser and Ellsworth, 2000.

* Use information from controlled sources (quarry blasts and shots) to estimate uncertainties of absolute locations and absolute origin times: [S16] Lin *et al.*, 2007; [S21] Lin *et al.*, 2007.

* Absolute locations for quarry seismicity: [S22] Lin *et al.*, 2006.

* Cut-and-paste method to source estimation: [S13] Zhu and Helmberger, 1996; [S27] Zhao and Helmberger, 1994.

* Use amplitude ratios between P and S waves to constrain the focal mechanisms: [S25] Hardebeck and Shearer, 2003.

* Moment tensor inversions with SEM: [S31] Liu *et al.*, 2004.

[7] Carl Tape, Qinya Liu, and Alessia Maggi *et al.*, 2009, Science, Adjoint tomography of the southern California crust. Date: 2016/11/27 Sun.

7.2 Adjoint tomography

Different results from different data sets: seismic reflection and industry well-log data to constrain the geometry and structure of major basins, receiver function data to estimate the depth to the Mohorovicic discontinuity, and local earthquake data to obtain the 3D background wave-speed structure.

Combine shear wave speed V_S and bulk sound speed V_B to compute compressional wave speed: $V_P^2 = (4/3)V_S^2 + V_B^2$.

An earthquake not used in the tomographic inversion or any future earthquake may be used to independently assess the misfit reduction of the inversion which use these data deriven from other earthquakes.

The approach is that of a minimization problem: (1) specification of an initial model that describes a set of earthquake source parameters and 3D variations in density, shear wave speed and bulk sound speed; (2) specification of a misfit function; (3) computation of the value of the misfit function for the initial model; (4) computation of the gradient and/or Hessian of the misfit function for the initial model; and (5) iterative minimization of the misfit function.

7.2.1 Misfit function

A given time window, or “measurement window”, is selected if there is a user-specified, quantifiable level of agreement between the observed and simulated seismograms.

For a single time window on a single seismogram, the travel-time misfit measure is:

$$F_i^T(\mathbf{m}) = \int_{-\infty}^{\infty} \frac{h_i(\omega)}{H_i} \Big[\frac{\Delta T_i(\omega, \mathbf{m})}{\sigma_i}\Big]^2 d\omega$$

where $\mathbf{m}$ is a model vector, $\Delta T_i(\omega, \mathbf{m}) = T_i^{obs}(\omega) - T_i^{syn}(\omega, \mathbf{m})$ is the frequency-dependent travel-time measurement associated with the ith window, σ_i is the estimated uncertainty associated with the travel-time measurement ($\sigma_i \geqslant \sigma_0$ the “water-level” minimum), and $h_i(\omega)$ is a frequency-domain window with associated normalization constant $H_i = \int_{-\infty}^{\infty} h_i(\omega) d\omega$ (the multi-taper method). If independent of frequency, $F_i^T(\mathbf{m}) = [\Delta T_i(\mathbf{m})/\sigma_i]^2$. For a single earthquake, the misfit function is:

$$F_s^T(\mathbf{m}) = \frac{1}{2} \frac{1}{N_s} \sum_{i=1}^{N_s} F_i^T(\mathbf{m})$$

where N_s denotes the total number of measurement windows for earthquake s. Overall misfit function is:

$$F^(\mathbf{m}) = \frac{1}{S} \sum_{s=1}^{S} F_s^T(\mathbf{m})$$

where S is the number of earthquakes.

7.3 Misfit analysis

Use the travel-time misfit measure within the tomographic inversion and the waveform misfit measure to assess the misfit reduction.

For a single time window on a single seismogram, the waveform misfit measure is:

$$F_i^W(\mathbf{m}) = \frac{\int_{-\infty}^{\infty} w_i(t)[d(t) - s(t, \mathbf{m})]^2 dt}{\left\{ \int_{-\infty}^{\infty} w_i(t)[d(t)]^2 dt \int_{-\infty}^{\infty} w_i(t)[s(t, \mathbf{m})]^2 dt \right\}^{1/2}}$$

where $d(t)$ denotes the recorded time series, $s(t, \mathbf{m})$ the simulated time series, $w_i(t)$ the ith time-domain window.

7.4 Earthquake source parameters

Four criteria, in order of importance, influenced selection of earthquakes for the tomographic inversion: (1) availability of good quality seismic waveforms for the period range of interest (must have at least 10 good stations); (2) availability of a relocated hypocenter (with origin time); (3) occurrence in a region with few other earthquake; (4) availability of a "reasonable" initial focal mechanism.

The dense coverage of stations in the vicinity of the earthquakes is important for epicenter estimation, as well as for depth and origin time.

8 Liu_2006_BSSA_Finite-frequency kernels[8]

8.1 Introduction

* Calculate sensitivity or Fréchet kernels: Marquering *et al.*, 1999 (surface-wave Green's functions); Zhao *et al.*, 2000 (normal modes); Dahlen *et al.*, 2000 & Hung *et al.*, 2000 & Zhou *et al.*, 2004 (asymptotic ray-based methods).
* Implement 3D travel-time ("banana-doughnut") kernels for compressional-wave tomography: Montelli *et al.*, 2004.
* *************************
* Obtain 3D finite-frequency sensitivity kernels for 3D reference models by calculating and storing 3D Green's functions: Zhao *et al.*, 2005.
* Obtain the gradient of a misfit function based on just a regular and an "adjoint" simulation for each earthquake: Tromp *et al.*, 2005.
* *************************
* Use spectral element method (SEM) on global or regional scales: Komatitsch and Tromp, 1999 & 2002a & 200b; Chaljub *et al.*, 2003; Komatitsch *et al.*, 2004.
* Implement the SEM on parallel computers: Komatitsch *et al.*, 2003.
* Calculate synthetic seismograms based on SEM: Komatitsch *et al.*, 2004.
* Paraxial absorbing equation: Clayton and Engquist, 1977; Quarteroni *et al.*, 1998.
* Perfectly matched layer (PML) methodology: Bérenger , 1994; Collino and Tsogka, 2001; Komatitsch and Tromp, 2003; Festa and Vilotte, 2005.
* The width of the 1st Fresnel zone is $\sqrt{\lambda L}$: Dahlen *et al.*, 2000.
* Los Angeles basin model: Hauksson, 2000 (background model); Süss and Shaw, 2003 (detailed).
* 3D source inversion technique: Liu *et al.*, 2004.
* Salton Trough model: Hauksson, 2000 (background model); Lovely *et al.*, 2006.
* P_{nl} wave train: Helmberger and Engen, 1980 (combination of the P_n and PL phases).
* Conjugate gradient approaches for 2D adjoint method: Tape *et al.*, 2006.

The main benefit of the adjoint approach is that the Fréchet derivatives of the misfit function may be obtained based on two 3D simulations for each earthquake. A disadvantage is the fact that the Hessian is unavailable, which leads to the use of iterative methods in the inversion problem.

[8] Qinya Liu and Jeroen Tromp, 2006, Bulletin of the Seismological Society of America, Finite-frequency kernels based on adjoint methods. Date: 2016/12/23 Fri.

8.2 Lagrange multiplier method

To minimize the least-squares waveform misfit function:

$$\chi = \frac{1}{2}\sum_r \int_0^T ||\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)||^2 dt$$

where $[0, T]$ denotes the time series of interest, $\mathbf{s}(\mathbf{x}_r, t)$ the synthetic and $\mathbf{d}(\mathbf{x}_r, t)$ the observed displacement at receiver $\mathbf{x}_r$ on time t. In practice, both $\mathbf{d}$ and $\mathbf{s}$ will be windowed, filtered, and possibly weighted.

An Earth model with volume Ω and outer free surface $\partial\Omega$. The synthetic $\mathbf{s}(\mathbf{x}, t)$ is determinde by:

$$\rho\partial_t^2\mathbf{s} - \nabla\cdot\mathbf{T} = \mathbf{f}$$

$$\mathbf{T} = \mathbf{c} : \nabla\mathbf{s}$$

where ρ density and $\mathbf{c}$ elastic tensor. The boundary condition and the initial conditions:

$$\hat{\mathbf{n}}\cdot\mathbf{T} = 0, \qquad \text{on } \partial\Omega$$

$$\mathbf{s}(\mathbf{x}, 0) = 0, \qquad \partial_t\mathbf{s}(\mathbf{x}, 0) = 0$$

where $\hat{\mathbf{n}}$ the unit outward normal. A simple point source at $\mathbf{x}_s$ in terms of the moment tensor $\mathbf{M}$ and source time function $S(t)$ as:

$$\mathbf{f} = -\mathbf{M}\cdot\nabla\delta(\mathbf{x} - \mathbf{x}_s)S(t)$$

Minimizing the misfit function when $\mathbf{s}$ satisfies wave equation implies

$$\chi = \frac{1}{2}\sum_r \int_0^T [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]^2 dt - \int_0^T \int_\Omega \lambda\cdot(\rho\partial_t^2\mathbf{s} - \nabla\cdot\mathbf{T} - \mathbf{f})d^3\mathbf{x}dt$$

where the vector Lagrange multiplier $\lambda(\mathbf{x}, t)$ remains to be determined. Using Hooke's law, upon integrating by parts (more detials refer to eq.9 of the original noted paper), because of the free surface boundary condition and the initial conditions,

$$\begin{aligned}
\delta\chi = &\int_0^T \int_\Omega \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]\delta(\mathbf{x} - \mathbf{x}_r)\cdot\delta\mathbf{s}(\mathbf{x}, t)d^3\mathbf{x}dt \\
&- \int_0^T \int_\Omega (\delta\rho\lambda\cdot\partial_t^2\mathbf{s} + \nabla\lambda : \delta\mathbf{c} : \nabla\mathbf{s} - \lambda\cdot\delta\mathbf{f})d^3\mathbf{x}dt - \int_0^T \int_\Omega [\rho\partial_t^2\lambda - \nabla\cdot(\mathbf{c} : \nabla\lambda)]\cdot\delta\mathbf{s}d^3\mathbf{x}dt \\
&- \int_\Omega [\rho(\lambda\cdot\partial_t\delta\mathbf{s} - \partial_t\lambda\cdot\delta\mathbf{s})]_T d^3\mathbf{x} - \int_0^T \int_{\partial\Omega} \hat{\mathbf{n}}\cdot(\mathbf{c} : \nabla\lambda)\cdot\delta\mathbf{s}d^2\mathbf{x}dt
\end{aligned}$$

where $[f]_T$ means $f(T)$.

If no model parameter perturbations $\delta\rho$, $\delta\mathbf{c}$ and $\delta\mathbf{f}$, and $\delta\chi$ vanish. In terms of $\delta\mathbf{s}$, λ statisfies

$$\rho\partial_t^2\lambda - \nabla\cdot(\mathbf{c} : \nabla\lambda) = \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]\delta(\mathbf{x} - \mathbf{x}_r)$$

with the free surface boundary condition and the end conditions:

$$\hat{\mathbf{n}}\cdot(\mathbf{c} : \nabla\lambda) = 0, \qquad \text{on } \partial\Omega$$

$$\lambda(\mathbf{x}, T) = 0, \qquad \partial_t\lambda(\mathbf{x}, T) = 0$$

Generally, λ is provided by the above equations, the variation reduces to

$$\delta\chi = -\int_0^T \int_\Omega (\delta\rho\lambda\cdot\partial_t^2\mathbf{s} + \nabla\lambda : \delta\mathbf{c} : \nabla\mathbf{s} - \lambda\cdot\delta\mathbf{f})d^3\mathbf{x}dt$$

Define the adjoint wave field $\mathbf{s}^\dagger$ in terms of λ by

$$\mathbf{s}^\dagger(\mathbf{x}, t) \equiv \lambda(\mathbf{x}, T - t)$$

Thus $\mathbf{s}^\dagger$ satisfies

$$\rho\partial_t^2\mathbf{s}^\dagger - \nabla\cdot\mathbf{T}^\dagger = \sum_r [\mathbf{s}(\mathbf{x}_r, T-t) - \mathbf{d}(\mathbf{x}_r, T-t)]\delta(\mathbf{x}-\mathbf{x}_r)$$

with the free surface boundary condition and the initial conditions:

$$\hat{\mathbf{n}}\cdot\mathbf{T}^\dagger = 0, \qquad \text{on } \partial\Omega$$

$$\mathbf{s}^\dagger(\mathbf{x}, 0) = 0, \qquad \partial_t\mathbf{s}^\dagger(\mathbf{x}, 0) = 0$$

where define the adjoint stress $\mathbf{T}^\dagger = \mathbf{c} : \nabla\mathbf{s}^\dagger$.

If no source perturbation $\delta\mathbf{f}$ (some changes refer to the original noted paper), the gradient of misfit function may be rewritten:

$$\delta\chi = \int_\Omega (\delta\rho K_\rho + \delta\mathbf{c} :: \mathbf{K}_c) d^3\mathbf{x}$$

where $\delta\mathbf{c} :: \mathbf{K}_c = \delta c_{ijkl} K_{c_{ijkl}}$ and difine the kernels

$$K_\rho(\mathbf{x}) = -\int_0^T \mathbf{s}^\dagger(\mathbf{x}, T-t)\cdot\partial_t^2\mathbf{s}(\mathbf{x}, t)dt$$

$$\mathbf{K}_c(\mathbf{x}) = -\int_0^T \nabla\mathbf{s}^\dagger(\mathbf{x}, T-t)\nabla\mathbf{s}(\mathbf{x}, t)dt$$

In an isotropic Earth model $c_{jklm} = (\kappa - {}^2\!/_3)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})$, where μ shear moduli and κ bulk moduli. Thus

$$\delta\mathbf{c} :: \mathbf{K}_c = \delta\ln\mu K_\mu + \delta\ln\kappa K_\kappa$$

where $\delta\ln\mu = {}^{\delta\mu}\!/_{\mu}$, $\delta\ln\kappa = {}^{\delta\kappa}\!/_{\kappa}$ and the isotropic kernels

$$K_\mu(\mathbf{x}) = -\int_0^T 2\mu(\mathbf{x})\mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t)dt$$

$$K_\kappa(\mathbf{x}) = -\int_0^T \kappa(\mathbf{x})[\nabla\cdot\mathbf{s}^\dagger(\mathbf{x}, T-t)][\nabla\cdot\mathbf{s}(\mathbf{x}, t)]dt$$

where the traceless strain deviator and its adjoint

$$\mathbf{D} = \frac{1}{2}[\nabla\mathbf{s} + (\nabla\mathbf{s})^T] - \frac{1}{3}(\nabla\cdot\mathbf{s})\mathbf{I}$$

$$\mathbf{D}^\dagger = \frac{1}{2}[\nabla\mathbf{s}^\dagger + (\nabla\mathbf{s}^\dagger)^T] - \frac{1}{3}(\nabla\cdot\mathbf{s}^\dagger)\mathbf{I}$$

where the superscript T denotes the transpose.

If in terms of ρ, shear wave speed β and compressional wave speed α,

$$K'_\rho = K_\rho + K_\kappa + K_\mu$$

$$K_\beta = 2\Big(K_\mu - \frac{4\mu}{3\kappa}K_\kappa\Big)$$

$$K_\alpha = 2\Big(\frac{\kappa + \frac{4}{3}\mu}{\kappa}\Big)K_\kappa$$

8.3 Absorbing boundaries

A regional Earth model has a boundary $\partial\Omega = \Sigma + \Gamma$, where Σ the free surface and Γ the artificial boundary. On Γ, absorbed energy based on the paraxial equation (Quarteroni *et al.*, 1998):

$$\hat{\mathbf{n}} \cdot \mathbf{T} = \rho[\alpha(\hat{\mathbf{n}}\hat{\mathbf{n}}) + \beta(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}})] \cdot \partial_t \mathbf{s} \equiv \mathbf{B} \cdot \partial_t \mathbf{s}, \qquad \text{on } \Gamma$$

In the original variation, substituting free surface boundary condition and the above absorbing boundary condition, upon integrating by parts,

$$\int_0^T \int_{\partial\Omega} \lambda \cdot [\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s})] - \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} d^2\mathbf{x} dt = -\int_0^T \int_\Sigma \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} d^2\mathbf{x} dt$$
$$+ \int_\Gamma [\lambda \cdot \mathbf{B} \cdot \delta\mathbf{s}]_T d^2\mathbf{x} - \int_0^T \int_\Gamma [\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) + \mathbf{B} \cdot \partial_t \lambda] \cdot \delta\mathbf{s} d^2\mathbf{x} dt$$

Thus, to vanish the Lagrange multiplier field, the free surface condition, the end condition and the absorbing boundary condition:

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) = 0, \qquad \text{on } \Sigma$$
$$\lambda(\mathbf{x}, T) = 0, \qquad \text{on } \Gamma$$
$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) = -\mathbf{B} \cdot \partial_t \lambda, \qquad \text{on } \Gamma$$

For the adjoint wave equation, the free surface boundary condition and the absorbing boundary condition:

$$\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = 0, \qquad \text{on } \Sigma$$
$$\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = \mathbf{B} \cdot \partial_t \mathbf{s}^\dagger, \qquad \text{on } \Gamma$$

which are same as for the regular wave equation.

8.4 Numerical implementation

If no attenuation, to reconstruct the forward wave field, backward in time from the displacement and velocity wave field at the end of the simulation. The backward wave equation is:

$$\rho\partial_t^2 \mathbf{s} = \nabla \cdot (\mathbf{c} : \nabla\mathbf{s}) + \mathbf{f}$$
$$\mathbf{s}(\mathbf{x}, T) \text{ and } \partial_t \mathbf{s}(\mathbf{x}, T) \text{ given}$$
$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\mathbf{s}) = 0, \text{ on } \partial\Omega$$

Technically, the only difference between solving the backward and forward wave equation is the change in the sign of the timestep parameter Δt.

For regional simulations, by saving the wave field on the absorbing boundaries at every timestep and the entire wave field at the end, reconstruct the forward wave field in reverse time by solving the backward wave equation, reinjecting the absorbed wave field as going along.

9 Tape_2007_GJI_Adjoint tomography 2D[9]

9.1 Introduction

* 3D Fréchet sensitivity kernels based on 1D reference model: Marquering *et al.*, 1999; Zhao *et al.*, 2000; Dahlen *et al.*, 2000.

[9] Carl Tape, Qinya Liu and Jeroen Tromp, 2007, Geophys. J. Int., Finite-frequency tomography using adjoint methods – Methodology and examples using membrane surface waves. Date: 2017/1/30 Mon.

* Seismic wave forward problem in complex media (SEM): Komatitsch and Vilotte, 1998; Komatitsch *et al.*, 2002; Capdeville *et al.*, 2003.
* *************************
* SEM 3D seismic wave propagation at regional and global scales: Komatitsch and Tromp, 1999; Komatitsch *et al.*, 2004; Komatitsch and Tromp, 2002a & 2002b.
* *************************
* Adjoint methods: Tarantola, 1984; Talagrand and Courtier, 1987.
* *************************
* Adjoint methods in exploration geophysics (2D): Tarantola, 1984; Gauthier *et al.*, 1986; Mora, 1987; Pratt *et al.*, 1998; Pratt, 1999.
* 3D ray tracing though 3D models to iteratively improve a global P-wave model: Bijwaard and Spakman, 2000.
* Fully finite difference method to compute traveltime misfit function gradients for 3D models of Los Angeles: Zhao *et al.*, 2005.
* A technique of stacking synthetic records that limits the number of forward simulations to one per event (per model iteration): Capdeville *et al.*, 2005.
* Tomographic inversion using finite-element method and adjoint approach within a conjugate gradient framework: Akcelik *et al.*, 2003.
* *************************
* Time-reversal imaging: Fink *et al.*, 1989; Fink, 1992 & 1997.
* *************************
* Classical tomography (compute model sensitivities for each measurement by constructing the gradient and Hessian of the misfit function): Woodhouse and Dziewonski, 1984; Ritsema *et al.*, 1999.
* *************************
* Membrane wave: Tanimoto, 1990; Peter *et al.*, 2006.
* Spherical spline basis functions to expand the fractional wave speed perturbations: Wang and Dahlen, 1995; Wang *et al.*, 1998.
* Crustal structure and seismicity distribution in southern California: Hauksson, 2000.
* Moho depth in southern California: Zhu and Kanamori, 2000.
* Calculate cheaply and rapidly banana-doughnut kernels for 1-D earth models: Dahlen *et al.*, 2000.
* Compute global finite-frequency kernels using normal modes for spherically symmetric models: Zhao and Jordan, 2006.
* Data weighting in waveform inversion: Takeuchi and Kobayashi, 2004.
* Add an explicit damping term to the misfit function to smooth the inversion: Akcelik *et al.*, 2002 & 2003.

* Conjugate gradient method: Fletcher and Reeves, 1964.
* Multiscale inversion method: Bunks *et al.*, 1995.

Seismic tomography based upon a 3-D reference model, 3-D numerical simulations depends largely on: (1) The accuracy and efficiency of the technique used to generate 3-D synthetic seismograms; (2) The efficiency of the inversion algorithm.

9.2 Inverse problem

Make a quadratic Taylor expansion of the misfit function $\chi(\mathbf{m} + \delta\mathbf{m})$,

$$\chi(\mathbf{m} + \delta\mathbf{m}) \approx \chi(\mathbf{m}) + \mathbf{g}(\mathbf{m})^T \delta\mathbf{m} + \frac{1}{2}\delta\mathbf{m}^T \mathbf{H}(\mathbf{m})\delta\mathbf{m}$$

where $\mathbf{m}$ a particular model, $\delta\mathbf{m}$ model corrections, and the gradient vector and the Hessian matrix are, respectively:

$$\mathbf{g}(\mathbf{m}) = \frac{\partial \chi}{\partial \mathbf{m}}\Big|_{\mathbf{m}}, \quad \mathbf{H}(\mathbf{m}) = \frac{\partial^2 \chi}{\partial \mathbf{m}\partial \mathbf{m}}\Big|_{\mathbf{m}}$$

The gradient with respect to $\delta\mathbf{m}$ is

$$\mathbf{g}(\mathbf{m} + \delta\mathbf{m}) \approx \mathbf{g}(\mathbf{m}) + \mathbf{H}(\mathbf{m})\delta\mathbf{m}$$

which can be set equal to zero to obtain the local minimum of misfit,

$$\mathbf{H}(\mathbf{m})\delta\mathbf{m} = -\mathbf{g}(\mathbf{m})$$

$$\delta\mathbf{m} = -\frac{\mathbf{g}(\mathbf{m})}{\mathbf{H}(\mathbf{m})}$$

If the gradient and (approximate) Hessian are both available, then the inverse approach is Newton method; if only the gradient is available, then it is gradient method.

9.3 Classical tomography

9.3.1 Theory

The traveltime misfit function may be

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{N}[T_i^{obs} - T_i(\mathbf{m})]^2$$

where T_i^{obs} and $T_i(\mathbf{m})$ the observed and predicted (based upon $\mathbf{m}$) traveltime for the ith source-receiver combination, and N the number of traveltime measurements. The variation is

$$\delta\chi = -\sum_{i=1}^{N}\Delta T_i \delta T_i$$

where δT_i the theoretical traveltime perturbation, and the traveltime anomaly:

$$\Delta T_i = T_i^{obs} - T_i(\mathbf{m})$$

where Δ and δ denote a differential measurement and a mathematical perturbation, respectively.

In ray-based tomography, the predicted traveltime anomaly δT_i along the ith ray path may be

$$\delta T_i = -\int_{ray_i} c^{-1}\delta \ln c ds$$

where fractional wave speed perturbations $\delta \ln c = {}^{\delta c}/_{c}$, and ds a segment of the ith ray. Taking into account finite-frequency effects, the traveltime anomaly for the ith source-receiver combination may be

$$\delta T_i = \int_V K_i \delta \ln c d^3\mathbf{x}$$

where $K_i(\mathbf{x})$ *'banana-doughnut', sensitivity, finite-frequency or Born kernels*.

For finite-frequency tomography,

$$\delta \chi = \int_V K \delta \ln c d^3\mathbf{x}$$

where the traveltime misfit kernel

$$K(\mathbf{x}) = -\sum_{i=1}^{N} \Delta T_i K_i(\mathbf{x})$$

Note that misfit kernels $K(\mathbf{x})$ depend upon the data, whereas the banana-doughnut kernels $K_i(\mathbf{x})$ are data-independent.

Choose a finite set of basis functions $B_k(\mathbf{x})$, $k = 1, 2, \ldots, M$ and expand fractional phase-speed perturbations,

$$\delta \ln c(\mathbf{x}) = \sum_{k=1}^{M} \delta m_k B_k(\mathbf{x})$$

where δm_k the perturbed model coefficients, determined in terms of $\mathbf{g}$ and $\mathbf{H}$ by the preceding relation.

Thus,

$$\delta T_i = \sum_{k=1}^{M} \delta m_k G_{ik}$$

where

$$G_{ik} \equiv \frac{\partial T_i}{\partial m_k}\Big|_{\mathbf{m}} = \begin{cases} -\displaystyle\int_{ray_i} c^{-1} B_k ds, & \text{for ray theory} \\ \displaystyle\int_V K_i B_k d^3\mathbf{x}, & \text{for finite-frequency tomography} \end{cases}$$

Besides, for finite-frequency tomography,

$$\delta \chi = \sum_{k=1}^{M} \delta m_k \int_V K B_k d^3\mathbf{x}$$

and

$$\delta \chi = \frac{\partial \chi}{\partial m} \cdot \delta \mathbf{m} = \mathbf{g} \cdot \delta \mathbf{m} = \sum_{k=1}^{M} g_k \delta m_k$$

deduce that

$$g_k = \frac{\partial \chi}{\partial m_k} = \int_V K B_k d^3\mathbf{x}$$

obtain

$$\begin{aligned} g_k &= -\sum_{i=1}^{N} \int_V K_i B_k d^3\mathbf{x} \Delta T_i \\ &= -\sum_{i=1}^{N} G_{ik} \Delta T_i \end{aligned}$$

In matrix notation,

$$\mathbf{g} = -\mathbf{G}^T\mathbf{d}$$

$$\mathbf{d} = (\Delta T_1, \Delta T_2, \dots, \Delta T_N)^T$$

As for ray-based tomography, same as finite-frequency tomography.

Because of

$$\frac{\partial \Delta T_i}{\partial m_{k'}} = -\frac{\partial T_i}{\partial m_{k'}} = -G_{ik'}$$

thus the Hessian $\mathbf{H}$

$$H_{kk'} = \frac{\partial^2 \chi}{\partial m_k \partial m_{k'}}\Big|_{\mathbf{m}} = \frac{\partial g_k}{\partial m_{k'}}\Big|_{\mathbf{m}} = \sum_{i=1}^{N}\Big(G_{ik}G_{ik'} - \Delta T_i \frac{\partial^2 T_i}{\partial m_k \partial m_{k'}}\Big|_{\mathbf{m}}\Big)$$

and the approximate Hessian $\tilde{\mathbf{H}}$

$$\tilde{H}_{kk'} \equiv \sum_{i=1}^{N} G_{ik}G_{ik'}$$

In matrix notation,

$$\tilde{\mathbf{H}} \equiv \mathbf{G}^T\mathbf{G}$$

If using the approximate Hessian instead of the exact one, then the inverse approach is a Gauss-Newton method.

Therefore, the model correction $\delta\mathbf{m}$ is determined by

$$\mathbf{G}^T\mathbf{G}\delta\mathbf{m} = \mathbf{G}^T\mathbf{d}$$

In general, $\tilde{\mathbf{H}}$ is not full rank. Introduce a damping matrix $\mathbf{D}$ (typically the norm, gradient, or second derivative of wave speed perturbations) and a damping parameter γ (generally determined by trading-off misfit of the solution against complexity of the model),

$$\tilde{\mathbf{H}}_\gamma = \mathbf{G}^T\mathbf{G} + \gamma^2\mathbf{D}$$

$$\delta\mathbf{m} = (\mathbf{G}^T\mathbf{G} + \gamma^2\mathbf{D})^{-1}\mathbf{G}^T\mathbf{d}$$

More details about how to add a regularization term to the misfit function refer to Appendix A of the original papaer. For non-linear inverse problems, an iterative Gauss-Newton method to minimize the misfit function.

9.3.2 Experimental set-up

2-D elastic wave equation for Membrane wave (traveling in the $x-y$ plane with a vertical z component of motion):

$$\rho\partial_t^2 s = \partial_x(\mu\partial_x s) + \partial_y(\mu\partial_y s) + f$$

where $s(x, y, t)$ the vertical component of displacement, $\rho(x, y)$ the density, $\mu(x, y)$ the shear modulus and the source

$$f(x, y, t) = h(t)\delta(x - x_s)\delta(y - y_s)$$

where $h(t)$ the source-time function and (x_s, y_s) the source location. A Gaussian form of the source-time function:

$$h(t) = -\frac{2\alpha^3}{\sqrt{\pi}}(t - t_s)e^{-\alpha^2(t-t_s)^2}$$

The relationship $\mu = \rho c^2$ and c is the membrane-wave phase-speed.

9.4 The gradient

For the 2-D case, the gradient of the misfit function is

$$g_k = \int_\Omega K B_k d^2\mathbf{x}$$

where K the misfit kernel.

9.4.1 Event kernels

The source for the adjoint wavefield for a particular event is (Tromp *et al.*, 2005, eq.57)

$$f^\dagger(x, y, t) = -\sum_{r=1}^{N_r} \Delta T_r \frac{1}{M_r} w_r(T - t)\partial_t s(x_r, y_r, T - t) \times \delta(x - x_r)\delta(y - y_r)$$

where r the receiver index, N_r the number of receivers, ΔT_r the cross-correlation traveltime measurement over a time window $w_r(t)$, $s(x, y, t)$ the forward wavefield, (x_r, y_r) the location of the receiver, T the length of the time-series, and M_r the normalization factor.

The adjoint source comprises time-reversed velocity seismograms, input at the location of the receivers and weighted by the traveltime measurement associated with each receiver.

For a given earthquake (event), the membrane event kernel:

$$K(x, y) = -2\mu(x, y) \int_0^T [\partial_x s^\dagger(x, y, T - t)\partial_x s(x, y, t) + \partial_y s^\dagger(x, y, T - t)\partial_y s(x, y, t)]dt$$

where $s^\dagger$ the adjoint wavefield given by the above adjoint source.

For a single receiver and a uniform model perturbation, the event kernel resembles a banana-doughnut kernel. The event kernel shows the region of the current model that gives rise to the discrepancy between the data and the synthetics.

To obtain a negative variation of the misfit function $\delta\chi$ to minimize the misfit, invoke a fast and positive structural perturbation where the kernel is negative, and/or a slow and negative structural perturbation where the kernel is positive.

9.4.2 Misfit kernels

Define the misfit kernel as a sum of event kernels for a particular model.

To remove spurious amplitudes in the vicinity of the sources and receivers, smooth the misfit kernel by convolving (in 2-D) the original misfit kernel with a Gaussian form:

$$G(x, y) = \frac{4}{\pi\Gamma^2} e^{-4(x^2+y^2)/\Gamma^2}$$

where Γ the scalelength of smoothing. The choice of Γ involves a degree of subjectivity, and it is feasible to take the value somewhat less than the wavelengths of the seismic waves.

The smoothing operation will tend to remove some subresolution features from the kernel.

9.4.3 Basis function

The basis functions embedded in the numerical method, using Lagrange polynomials for the SEM, refer to the Section 5.3 of the original paper.

9.5 Optimization

9.5.1 Conjugate gradient algorithm

Given an initial model $\mathbf{m}^0$, calculate $\chi(\mathbf{m}^0), \mathbf{g}^0 = {}^{\partial\chi}/_{\partial\mathbf{m}}(\mathbf{m}^0)$, and set the initial search direction $\mathbf{p}^0 = -\mathbf{g}^0$. If $||\mathbf{p}^0|| < \epsilon$, then $\mathbf{m}^0$ is the desired model; otherwise:

(i) Perform a line search to obtain the scalar v_k that minimizes the function $\tilde{\chi}^k(v)$, where $\tilde{\chi}^k(v) = \chi(\mathbf{m}^k + v\mathbf{p}^k)$ and $\tilde{g}^k(v) = {}^{\partial\tilde{\chi}^k}/_{\partial v} = \mathbf{g}(\mathbf{m}^k + v\mathbf{p}^k)\cdot\mathbf{p}^k$:

- Choose a test parameter $v_t^k = -{}^{2\tilde{\chi}^k(0)}/_{\tilde{g}^k(0)}$;
- Calculate the test model $\mathbf{m}_t^k = \mathbf{m}^k + v_t^k\mathbf{p}^k$, $\chi(\mathbf{m}_t^k)$, $\mathbf{g}(\mathbf{m}_t^k)$, $\tilde{\chi}^k(v_t^k)$ and $\tilde{g}^k(v_t^k)$;
- Interpolate the function $\tilde{\chi}^k(v)$ by a quadratic or cubic polynomial (resolve a quadratic or cubic polynomial $\tilde{\chi}^k(v)$ according to the two misfits $\tilde{\chi}^k(0)$, $\tilde{\chi}^k(v_t^k)$, the gradient(s) $\tilde{g}^k(0)$, not or and $\tilde{g}^k(v_t^k)$) and obtain the v^k that gives the minimum of this polynomial (more details refer to Appendix B2 of the original paper).

(ii) Update the model: $\mathbf{m}^{k+1} = \mathbf{m}^k + v^k\mathbf{p}^k$, then calculate $\mathbf{g}^{k+1} = {}^{\partial\chi}/_{\partial\mathbf{m}}(\mathbf{m}^{k+1})$.

(iii) Update the conjugate gradient search direction: $\mathbf{p}^{k+1} = -\mathbf{g}^{k+1} + \beta^{k+1}\mathbf{p}^k$, where $\beta^{k+1} = \mathbf{g}^{k+1}\cdot {}^{(\mathbf{g}^{k+1}-\mathbf{g}^k)}/_{(\mathbf{g}^k\cdot\mathbf{g}^k)}$.

(iv) If $||\mathbf{p}^{k+1}|| < \epsilon$, then $\mathbf{m}^{k+1}$ is the desired model; otherwise replace k with $k+1$ and restart from (i).

A detailed cycle of the conjugate gradient algorithm for the adjoint tomography refer to the Fig.11 of the original paper.

Entrapment into local minima is common in the conjugate gradient method, and it may be avoided by using multiscale methods (Bunks *et al.*, 1995), and alternatively by starting at longer periods and gradually moving to shorter periods.

9.6 Source, structure and joint inversions

9.6.1 Source inversion

A perturbation of the point source may be:

$$\delta f(x,y,t) = -\dot{h}(t)\delta t_s\delta(x-x_s)\delta(y-y_s) + h(t)(\delta x_s\partial_{x_s} + \delta y_s\partial_{y_s})[\delta(x-x_s)\delta(y-y_s)]$$

where δt_s a perturbation in the origin time, $(\delta x_s, \delta y_s)$ a perturbation in the source location.

Change in misfit due to a change in the point source is

$$\delta\chi = \int_0^T \int_\Omega \delta f(x,y,t)s^\dagger(x,y,T-t)dxdydt$$

where $s^\dagger$ the adjoint wavefield, whose sources are injected at the receivers, just same as in the case of the previous structure inversions. Thus,

$$\delta\chi = -\delta t_s\int_0^T \dot{h}(t)s^\dagger(x_s,y_s,T-t)dt + (\delta x_s\partial_{x_s} + \delta y_s\partial_{y_s})\int_0^T h(t)s^\dagger(x_s,y_s,T-t)dt$$

$$\delta\chi = \mathbf{g}\cdot\delta\mathbf{m}$$

where

$$\delta\mathbf{m} = \left[\frac{\delta x_s}{\lambda}, \frac{\delta y_s}{\lambda}, \frac{\delta t_s}{\tau}\right]^T = \left[\frac{x_s^k - x_s^0}{\lambda}, \frac{y_s^k - y_s^0}{\lambda}, \frac{t_s^k - t_s^0}{\tau}\right]^T$$

$$\mathbf{g} = \left[\lambda\int_0^T h(t)\partial_{x_s}s^\dagger(x_s,y_s,T-t)dt, \lambda\int_0^T h(t)\partial_{y_s}s^\dagger(x_s,y_s,T-t)dt, -\tau\int_0^T \dot{h}(t)s^\dagger(x_s,y_s,T-t)dt\right]$$

where τ the reference period, $\lambda = c\tau$ the reference wavelength and c the reference phase speed.

9.6.2 Joint inversions

The model vector for the joint inversion is $\delta\mathbf{m} = [\delta\mathbf{m}_{str}; \delta\mathbf{m}_{src}]$ with dimension $N_{structure}+3N_{event}$. The gradient is

$$\mathbf{g}^k = [F\mathbf{g}^k_{str}; \mathbf{g}^k_{src}]$$

$$F = \frac{||\mathbf{g}^0_{src}||_2}{||\mathbf{g}^0_{str}||_2}$$

where $||\cdot||_2$ the L2-norm of the enclosed vector.

9.7 Discussion

9.7.1 Three kernel types

Banana-doughnut kernels: a phase-specific (e.g. P) kernel for an individual source-receiver combination, not incorporate the measurement; Event kernels: a sum of individual banana-doughnut kernels, weighted by its corresponding measurement; Misfit kernel: the sum of event kernels, a graphical representation of the gradient of the misfit function.

Use the banana-doughnut kernels in classical tomography and the misfit kernels in adjoint tomography.

10 Bozdag_2011_GJI_Misfit functions for FWI[10]

10.1 Introduction

* Full waveform inversions in local and regional studies: Chen *et al.*, 2007b; Fichtner *et al.*., 2009; Tape *et al.*, 2009.

* A global tomography approach in a synthetic experiment based on a source stacking technique: Capdeville *et al.*, 2005.

* ************************

* Ray-based tomography: Zhou, 1996 & Boschi and Dziewonski, 2000 (using body-wave phases); Trampert and Woodhouse, 1995 & Ekstrom *et al.*, 1997 (using surface waves).

* Integration of different data sets to increase resolution in ray-based tomography: Su *et al.*, 1994; Masters *et al.*, 1996; Ritsema *et al.*, 1999; Megnin and Romanowicz, 2000; Gu *et al.*, 2001.

* Finite-frequency tomography to improve resolution: Montelli *et al.*, 2004; Sigloch *et al.*, 2008; Boschi *et al.*, 2007.

* Construct global models based on energy wave packets using asymptotic finite-frequency kernels: Li and Romanowicz, 1996; Megnin and Romanowicz, 2000; Gung and Romanowicz, 2004.

* ************************

* Solve the wave euqation numerically in realistic 3-D earth models: Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999; Capdeville *et al.*, 2003.

* Compute Green's functions in 3-D models to compute Fréchet derivatives: Zhao *et al.*, 2005.

[10] Ebru Bozdag, Jeannot Trampert and Jeroen Tromp, 2011, Geophys. J. Int., Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Date: 2017/4/3 Mon.

* Adjoint techniques: Tarantola, 1984 & 1988; Fink, 1997; Talagrand and Courtier, 1987; Crase *et al.*, 1990; Pratt, 1999; Akcelik *et al.*, 2003.
* Combine 3-D simulations with adjoint techniques to compute Fréchet derivatives: Tromp *et al.*, 2005.
* Compare the scattering integral method (computing and storing 3-D Green's functions) with adjoint methods: Chen *et al.*, 2007a.
* ************************
* Common misfit functions based on: Luo and Schuster, 1991 & Marquering *et al.*, 1999 & Dahlen *et al.*, 2000 & Zhao *et al.*, 2000 (cross-correlation traveltime measurements); Dahlen and Baig, 2002 & Ritsema *et al.*, 2002 (relative amplitude variations); Tarantola, 1984 & Tarantola, 1988 & Nolet, 1987 (waveform differences).
* Adjoint sensitivity kernels based on cross-correlation traveltime measurements: Liu and Tromp, 2006 & 2008.
* Automated phase-picking algorithms: Maggi *et al.*, 2009.
* Multitaper measurements: Zhou *et al.*, 2004.
* Regional example of seismic tomography based on frequency-dependent traveltime measurements and multitaper measurements using CG method with adjoint kernels: Tape *et al.*, 2009.
* Generalized seismological data functionals (GSDF) for frequency-dependent measurements: Gee and Jordan, 1992.
* Time-frequency analysis separating phase and amplitude information: Fichtner *et al.*, 2008.
* ************************
* Instantaneous phases to increase resolution in exploration seismics: Taner *et al.*, 1979; Perz *et al.*, 2004; Barnes, 2007.
* ************************
* Spectral-element method (SEM): Komatitsch and Tromp 2002a & 2002b.
* PREM: Dziewonski and Anderson, 1981.
* 3-D mantle model S20RTS: Ritsema *et al.*, 1999.
* 3-D crustal model Crust2.0: Bassin *et al.*, 2000.
* 3-D Q model: Dalton *et al.*, 2008.

BACKGROUND: In classical seismic tomography, the usable amount of data is often restricted because of approximations to the wave equation. 3-D numerical simulations of wave propagation provide new opportunities for increasing the amount of usable data in seismograms by choosing appropriate misfit functions.

DEFINITIOIN: "Full waveform inversion" is a technique which combines 3-D numerical wave simulations as a forward theory with Fréchet kernels computed in 3-D background models, to fit complete three-component seismograms.

WHY: The waveform misfit is easily applied to whole seismograms, but it favours high-amplitude phases in a wave train containing multiple phases with different amplitudes. Thus, to extract optimal information, phases should be selected as in traveltime measurements, or seismograms should be cut into wave packages with appropriate weightings (e.g. Li and Romanowicz, 1996).

Waveform differences can be highly non-linear with respect to the model.

10.2 Misfit functions and adjoint sources

10.2.1 Adjoint kernels

WORKFLOW: In seismic waveform tomography, we extract information from a set of observed seismograms on model parameters describing Earth's interior. Model parameters are updated by minimizing a chosen misfit function between observed and synthetic data. In adjoint tomography, the gradient of the misfit function can be computed through the interaction of a forward wavefield with its adjoint wavefield, which is generated by the back-propagation of measurements made on data. The non-linear inverse problem is then solved iteratively based on a gradient method. Define a generic wavefor misfit function:

$$\chi(\mathbf{m}) = \sum_{r=1}^{N} \int_0^T g(\mathbf{x}_r, t, \mathbf{m}) dt$$

where N the number of receivers, $g(\mathbf{x}_r, t, \mathbf{m})$ any kind of misfit at receiver position $\mathbf{x}_r$ with model parameters $\mathbf{m}$. Its gradient is

$$\delta\chi = \sum_{r=1}^{N} \int_0^T \partial_{\mathbf{s}} g(\mathbf{x}_r, t, \mathbf{m}) \cdot \delta\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) dt$$

where $\delta\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})$ displacement perturbations due to model perturbations $\delta\mathbf{m}$. Using the Born approximation and the reciprocity of the Green's function, defines the adjoint wavefield $\mathbf{s}^\dagger$ and the adjoint source $\mathbf{f}^\dagger$:

$$s_k^\dagger(\mathbf{x}', t') = \int_0^{t'} \int_V G_{ki}(\mathbf{x}', \mathbf{x}_r; t' - t) f_i^\dagger(\mathbf{x}, t) d^3\mathbf{x} dt$$

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N} \partial_{s_i} g(\mathbf{x}_r, T - t, \mathbf{m}) \delta(\mathbf{x} - \mathbf{x}_r)$$

The sensitivity kernels are the Fréchet derivatives with respect to the corresponding model parameter.

Adjoint kernels depend on the adjoint wavefield, which is generated by the adjoint source. And the adjoint source depends on the pre-defined misfit function for specific observables.

10.2.2 Hilbert transform

An analytic signal $\tilde{f}(t)$ is constructed from a real signal $f(t)$ and its Hilbert transform $\mathcal{H}\{f(t)\}$:

$$\tilde{f}(t) = f(t) - i\mathcal{H}\{f(t)\}$$

$$\mathcal{H}\{f(t)\} = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(\tau)}{t - \tau} d\tau$$

where P the Cauchy principal value. The analytic signal can be written as

$$\tilde{f}(t) = E(t) e^{i\phi(t)}$$

where the instantaneous phase $\phi(t)$ and the instantaneous amplitude $E(t)$ are respectively:

$$\phi(t) = \arctan \frac{\mathcal{I}\{\tilde{f}(t)\}}{\mathcal{R}\{\tilde{f}(t)\}}$$

$$E(t) = \sqrt{\mathcal{R}\{\tilde{f}(t)\}^2 + \mathcal{I}\{\tilde{f}(t)\}^2}$$

10.2.3 Instantaneous phase misfits

Define the squared instantaneous phase misfit:

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T [\phi_r^{obs}(t) - \phi_r(t,\mathbf{m})]^2 dt$$

where ϕ_r the instantaneous phase of a specific component recorded at receiver r. Its gradient is

$$\delta\chi = -\sum_{r=1}^{N}\int_0^T (\phi_r^{obs} - \phi_r)\delta\phi_r dt$$

Assume that $\tilde{s}_r$ is the analytic signal corresponding to the synthetic seismogram s_r, define ϕ_r as

$$\phi_r = \arctan\frac{\mathcal{I}(\tilde{s}_r)}{\mathcal{R}(\tilde{s}_r)}$$

so the perturbation

$$\begin{aligned}\delta\phi_r &= \delta\Big[\tfrac{\mathcal{I}(\tilde{s}_r)}{\mathcal{R}(\tilde{s}_r)}\Big]/\Big\{1+\Big[\tfrac{\mathcal{I}(\tilde{s}_r)}{\mathcal{R}(\tilde{s}_r)}\Big]^2\Big\}\\ &= \frac{(\mathcal{H}s_r)\delta s_r - s_r\delta(\mathcal{H}s_r)}{s_r^2 + (\mathcal{H}s_r)^2} \quad \text{(some problem on derivation)}\\ &= \frac{(\mathcal{H}s_r)\delta s_r - s_r\delta(\mathcal{H}s_r)}{E_r^2}\end{aligned}$$

and the gradient

$$\begin{aligned}\delta\chi &= -\sum_{r=1}^{N}\int_0^T (\phi_r^{obs} - \phi_r)\Big[\frac{(\mathcal{H}s_r)\delta s_r}{E_r^2} - \frac{s_r\delta(\mathcal{H}s_r)}{E_r^2}\Big]dt\\ &= -\sum_{r=1}^{N}\int_0^T \Big[(\phi_r^{obs} - \phi_r)\frac{(\mathcal{H}s_r)}{E_r^2}\delta s_r + \mathcal{H}\Big\{(\phi_r^{obs} - \phi_r)\frac{s_r}{E_r^2}\Big\}\delta s_r\Big]dt\end{aligned}$$

Then the adjoint source

$$\begin{aligned}f_i^\dagger(\mathbf{x},t) = -\sum_{r=1}^{N}\Big[&[\phi_i^{obs}(\mathbf{x}_r,T-t) - \phi_i(\mathbf{x}_r,T-t,\mathbf{m})]\frac{w_r(T-t)\mathcal{H}\{s_i(\mathbf{x}_r,T-t,\mathbf{m})\}}{E_i(\mathbf{x}_r,T-t,\mathbf{m})^2}\\ &+\mathcal{H}\Big\{[\phi_i^{obs}(\mathbf{x}_r,T-t) - \phi_i(\mathbf{x}_r,T-t,\mathbf{m})]\frac{w_r(T-t)s_i(\mathbf{x}_r,T-t,\mathbf{m})}{E_i(\mathbf{x}_r,T-t,\mathbf{m})^2}\Big\}\Big]\delta(\mathbf{x}-\mathbf{x}_r)\end{aligned}$$

where w_r the weighting function, generically defined as $1/E_i^2$.

10.2.4 Envelope misfits

Define the squared logarithmic envelope misfit:

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T \Big[\ln\frac{E_r^{obs}(t)}{E_r(t,\mathbf{m})}\Big]^2 dt$$

where E_r the envelope of a specific component recorded at receiver r. Its gradient is

$$\delta\chi = -\sum_{r=1}^{N}\int_0^T \ln\Big(\frac{E_r^{obs}}{E_r}\Big)\frac{1}{E_r}\delta E_r dt$$

Similarly, define E_r as

$$E_r = \sqrt{\mathcal{R}(\tilde{s}_r)^2 + \mathcal{I}(\tilde{s}_r)^2}$$

so the perturbation

$$\delta E_r = \frac{s_r\delta s_r + (\mathcal{H}s_r)\delta(\mathcal{H}s_r)}{\sqrt{s_r^2 + (\mathcal{H}s_r)^2}}$$

and the gradient

$$\begin{aligned}\delta\chi &= -\sum_{r=1}^{N}\int_0^T \ln\Big(\frac{E_r^{obs}}{E_r}\Big)\Big[\frac{s_r\delta s_r}{E_r^2} + \frac{(\mathcal{H}s_r)\delta(\mathcal{H}s_r)}{E_r^2}\Big]dt \\ &= -\sum_{r=1}^{N}\int_0^T \Big[\ln\Big(\frac{E_r^{obs}}{E_r}\Big)\frac{s_r}{E_r^2}\delta s_r - \mathcal{H}\Big\{\ln\Big(\frac{E_r^{obs}}{E_r}\Big)\frac{(\mathcal{H}s_r)}{E_r^2}\Big\}\delta s_r\Big]dt\end{aligned}$$

Then the adjoint source

$$\begin{aligned}f_i^{\dagger}(\mathbf{x},t) &= -\sum_{r=1}^{N}\Big[\ln\Big[\frac{E_r^{obs}(\mathbf{x}_r,t)}{E_r(\mathbf{x}_r,t,\mathbf{m})}\Big]\frac{w_r(t)s_i(\mathbf{x}_r,T-t,\mathbf{m})}{E_i(\mathbf{x}_r,T-t,\mathbf{m})^2} \\ &= -\mathcal{H}\Big\{\ln\Big[\frac{E_r^{obs}(\mathbf{x}_r,t)}{E_r(\mathbf{x}_r,t,\mathbf{m})}\Big]\frac{w_r(t)\mathcal{H}\{s_i(\mathbf{x}_r,T-t,\mathbf{m})\}}{E_i(\mathbf{x}_r,T-t,\mathbf{m})^2}\Big\}\Big]\delta(\mathbf{x}-\mathbf{x}_r)\end{aligned}$$

where w_r the weighting function.

10.2.5 Waveform misfits

The classical misfit function is defined as

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T ||\mathbf{d}(\mathbf{x}_r,t) - \mathbf{s}(\mathbf{x}_r,t,\mathbf{m})||^2 dt$$

where **d** and **s** the observed and synthetic waveforms, respectively. Its gradient is

$$\delta\chi = -\sum_{r=1}^{N}\int_0^T [\mathbf{d}(\mathbf{x}_r,t) - \mathbf{s}(\mathbf{x}_r,t,\mathbf{m})]\delta\mathbf{s}(\mathbf{x}_r,t,\mathbf{m})dt$$

And the adjoint source is

$$f_i^{\dagger}(\mathbf{x},t) = -\sum_{r=1}^{N}\frac{1}{M_r}[d_i(\mathbf{x}_r,T-t) - s_i(\mathbf{x}_r,T-t,\mathbf{m})]w_r(T-t)\delta(\mathbf{x}-\mathbf{x}_r)$$

where w_r the time window function, and the normalization term $M_r = \int_0^T w_r(t)d_i^2(x_r,t)dt$.

10.2.6 Traveltime misfits

The squared traveltime misfit is

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}[T_r^{obs} - T_r(\mathbf{m})]^2$$

where T_r the traveltime of a selected phase at receiver r. Its gradient is

$$\delta\chi = -\sum_{r=1}^{N}[T_r^{obs} - T_r(\mathbf{m})]\delta T_r$$

If traveltime differences are measured by cross-correlation, the perturbation

$$\delta T_r = \frac{1}{N_r}\int_0^T w_r(t)\partial_t s_i(\mathbf{x}_r, t, \mathbf{m})\delta s_i(\mathbf{x}_r, t, \mathbf{m})dt$$

$$N_r = \int_0^N w_r(t)s_i(\mathbf{x}_r, t, \mathbf{m})\partial_t^2 s_i(\mathbf{x}_r, t, \mathbf{m})dt$$

where w_r the time window function which isolates a specific phase, and the adjoint source

$$f_i^\dagger(\mathbf{x}, t) = -\sum_{r=1}^{N}[T_r^{obs} - T_r(\mathbf{m})]\frac{1}{N_r}w_r(T-t)\partial_t s_i(\mathbf{x}_r, T-t, \mathbf{m})\delta(\mathbf{x}-\mathbf{x}_r)$$

10.2.7 Amplitude misfits

The amplitude misfit is

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}\Big[\ln\frac{A_r^{obs}}{A_r(\mathbf{m})}\Big]^2$$

where the amplitude $A_r = \sqrt{1/(t_2-t_1)\int_{t_1}^{t_2} s_r^2(t)dt}$ (Dahlen and Baig, 2002) at station r. Its gradient is

$$\delta\chi = -\sum_{r=1}^{N}\ln\Big[\frac{A_r^{obs}}{A_r(\mathbf{m})}\Big]\delta\ln A_r$$

$$\delta\ln A_r = \frac{1}{M_r}\int_0^T w_r(t)s_i(\mathbf{x}_r, t, \mathbf{m})\delta s_i(\mathbf{x}_r, t, \mathbf{m})dt$$

where w_r the time window function, and the normalization factor $M_r = \int_0^T w_r(t)s_i^2(\mathbf{x}_r, t, \mathbf{m})dt$. And the adjoint source

$$f_i^\dagger(\mathbf{x}, t) = -\sum_{r=1}^{N}\ln\Big[\frac{A_r^{obs}}{A_r(\mathbf{m})}\Big]\frac{1}{M_r}w_r(T-t)s_i(\mathbf{x}_r, T-t, \mathbf{m})\delta(\mathbf{x}-\mathbf{x}_r)$$

10.2.8 Attenuation kernels

Amplitudes or envelopes of seismograms are also very sensitive to variations in anelastic structure. Express the gradient of the misfit function

$$\delta\chi = \int_v K_\mu^Q(\mathbf{x})\delta Q_\mu^{-1}(\mathbf{x})d^3\mathbf{x}$$

where Q_κ^{-1} is ignored. The frequency-dependent shear modulus is (Liu *et al.*, 1976)

$$\mu(\omega) = \mu(\omega_0)\left[1 + \frac{2}{\pi}Q_\mu^{-1}\ln\frac{|\omega|}{\omega_0} - i\mathrm{sgn}(\omega)Q_\mu^{-1}\right]$$

where ω_0 a reference angular frequency, and the change (Tromp *et al.*, 2005)

$$\delta\mu(\omega) = \mu(\omega_0)\left[\frac{2}{\pi}\ln\frac{|\omega|}{\omega_0} - i\mathrm{sgn}(\omega)\right]\delta Q_\mu^{-1}$$

According to the Fourier transformed Born approximation, the anelastic adjoint wavefield is

$$\tilde{f}_i^\dagger(\mathbf{x}, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{2}{\pi}\ln\frac{|\omega|}{\omega_0} - i\mathrm{sgn}(\omega)\right]f_i^\dagger(\mathbf{x}, \omega)e^{i\omega t}d\omega$$

where $f_i^\dagger(\mathbf{x}, \omega)$ the Fourier transform of the regular elastic adjoint source.

10.3 Discussion

Waveform measurement (WF) favours the highest amplitude parts of seismograms.

The drawback of traveltime (TT) and amplitude (AMP) measurements is that they need waveforms to be similar in shape and require isolating seismic phases from seismograms (need to pick every available phase); The major disadvantage of WF comes from mixing phase and amplitude information in a single observable and is highly non-linear with respect to Earth's structure.

The advantages of instantaneous phase (IP) and envelope (ENV) measurements are less data processing and easier implementation.

To avoid cycle skip problems in phase speed measurements, use long-period waveforms first, gradually increase the frequency content of data in subsequent iterations in the inversion.

11 Moghaddam_2013_Geophy_Stochastic gradient method[11]

11.1 Introduction

* FWI on a large scale: Virieux and Operto, 2009; Kapoor *et al.*, 2010; Vigh *et al.*, 2010.
* Conventional FWI: Tarantola, 1984 & 1986; Mora, 1987; Crase *et al.*, 1990; ...
* Source encoding technique: Krebs *et al.*2009; Li and Herrmann, 2010; Moghaddam and Herrmann, 2010; van Leeuwen *et al.*, 2011; Haber *et al.*, 2012; Li *et al.*, 2012.
* *************************
* Stochastic optimization method: Goldberg, 1989; Spall, 1992.
* *************************
* Marmousi model: Bourgeois *et al.*, 1991.
* *************************
* The adjoint-state method to avoid the computation of sensitivity matrix: Lions and Magenes, 1972; Lailly, 1983; Tarantola, 1984; Giles *et al.*, 2003; Plessix, 2006; Virieux and Operto, 2009.

[11] Peyman P. Moghaddam, Henk Keers, Felix J. Herrmann, *et al.*, 2013, Geophysics, A new optimizatioin approach for source-encoding full-waveform inversion. Date: 2017/6/11 Sun.

* The limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method: Byrd *et al.*, 1995; Mulder and Plessix, 2004; Nocedal and Wright, 2006; Plessix, 2004.
* Preconditioned conjugate gradient method: Ravaut *et al.*, 2004.
* Gauss-Newton method: Virieux and Operto, 2009.
* The online LBFGS (oLBFGS): Schraudolph *et al.*, 2007; Yu *et al.*, 2010.
* Stochastic gradient descent: Schraudolph *et al.*, 2007; Sunehag *et al.*, 2009.

The misfit function, and therefore also its gradient, for source-encoding waveform inversion is an unbiased random estimation of the misfit function used in conventional waveform inversion.

Main drawbacks of FWI: the requirement to have an accurate initial model; and expensive computational cost.

Source encoding uses a linear combinations of all shots, with random weights assigned to each shot.

11.2 Stochastic optimization

Stochastic gradient descent Stochastic gradient descent is:

$$\sigma_{k+1} = \sigma_k - \eta_k \nabla J(\sigma_k, \mathbf{w}_k)$$

where k the iteration number, η_k the step length, J the misfit function, σ_k the model at iteration k and $\mathbf{w}_k$ the current randomized weight.

Stochastic LBFGS Each step of the LBFGS algorithm takes:

$$\sigma_{k+1} = \sigma_k - \eta_k \mathbf{H}_k \nabla J(\sigma_k, \mathbf{w}_k)$$

where the inverse Hessian matrix $\mathbf{H}_k$ updated in each iteration by (refer to the last second formula of Wikipedia page):

$$\mathbf{H}_{k+1} = \mathbf{V}_k^T \mathbf{H}_k \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T$$

with $\rho_k = {}^1/_{\mathbf{y}_k^T \mathbf{s}_k}$, $\mathbf{V}_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T$ and $\mathbf{s}_k = \sigma_{k+1} - \sigma_k$, $\mathbf{y}_k = \nabla J(\sigma_{k+1}, \mathbf{w}_k) - \nabla J(\sigma_k, \mathbf{w}_k)$. Note that for construction of $\mathbf{y}_k$, take the same random weighting $\mathbf{w}_k$ for the current gradient at $k+1$ and the previous one at k.

The LBFGS routine is carried out in two steps. First, the latest m iterations are calculated. Second, the routine updates the LBFGS direction as the following:

1: $\mathbf{q} \leftarrow \nabla J(\mathbf{m}_k, \mathbf{w}_k)$
2: $\mathbf{H}_k^0 \leftarrow {}^{(\mathbf{y}_k^T \mathbf{s}_k)}/_{(\mathbf{y}_k^T \mathbf{y}_k)}$
3: FOR $i = k$ to $k - m + 1$
4: $\quad \alpha_i \leftarrow \rho_i \mathbf{s}_i^T \mathbf{q}$
5: $\quad \mathbf{q} \leftarrow \mathbf{q} - \alpha_i \mathbf{y}_i$
6: END FOR
7: $\mathbf{r} \leftarrow \mathbf{H}_k^0 \mathbf{q}$
8: FOR $i = k - m + 1$ to k
9: $\quad \beta \leftarrow \rho_i \mathbf{y}_i^T \mathbf{r}$
10: $\quad \mathbf{r} \leftarrow \mathbf{r} + \mathbf{s}_i(\alpha_i - \beta)$
11: END FOR
12: Stop with $\mathbf{r} = \mathbf{H}_{k+1} \nabla J(\mathbf{m}_{k+1}, \mathbf{w}_{k+1})$

Stochastic oLBFGS For better convergence, the oLBFGS method uses $\mathbf{y}_k = \nabla J(\mathbf{m}_{k+1}, \mathbf{w}_k) - \nabla J(\mathbf{m}_k, \mathbf{w}_k) + \lambda \mathbf{s}_k$. And the step $\mathbf{r} \leftarrow \mathbf{H}_k^0 \mathbf{q}$ in the above procedures is replaced by:

$$\mathbf{r} = \frac{\mathbf{q}}{\min(k,m)} \sum_{i=1}^{\min(k,m)} \frac{\mathbf{s}_{k-i}^T \mathbf{y}_{k-i}}{\mathbf{y}_{k-i}^T \mathbf{y}_{k-i}}$$

where we can set $\lambda = 0.1 \cdot {}^{||J(\mathbf{m}_0, \mathbf{w}_0)||_2^2}/_{||\mathbf{m}_0||_2^2}$.

Integrated stochastic gradient descent To accelerate the convergence, in the integrated stochastic gradient descent (iSGD) method, the iteration step takes:

$$\sigma_{k+1} = \sigma_k - \eta_k \overline{\nabla J(\sigma_k)}$$

$$\overline{\nabla J(\sigma_k)} = \frac{\sum_{i=k-m}^{k} e^{\alpha(i-k)} \nabla J(\sigma_i, \mathbf{w}_i)}{\sum_{i=k-m}^{k} e^{\alpha(i-k)}}$$

where we can set $m = 10$.

12 Louboutin_2017_EAGE_Gradient sampling algorithm[12]

12.1 Introduction

* Wavefield reconstruction inversion (WRI) where both the velocity model and wavefields are unknown: van Leeuwen *et al.*, 2014; van Leeuwen and Herrmann, 2015.

12.2 Gradient sampling for FWI

12.2.1 Gradient sampling algorithm

By working with local neighborhoods instead with a single model, the algorithm is able to reap global information on the objective from local gradient at small cost.

Minimize the objective function $\Phi(\mathbf{x})$ with respect to $\mathbf{x} \in R^N$ by

- sampling $N+1$ vectors $\mathbf{x}_{ki}$ in a ball $B_{\epsilon_k}(\mathbf{x}_k)$ defined as all $\mathbf{x}_{ki}$ such that $||\mathbf{x}_k - \mathbf{x}_{ki}||_2^2 < \epsilon_k$, where ϵ_k is the maximum distance between the current estimate and a sampled vector;
- calculating gradients for each sample, i.e., $\mathbf{g}_{ki} = \nabla\Phi(\mathbf{x}_{ki})$;
- computing descent directions as a weighted sum over all sampled gradients, i.e., $\mathbf{g}_k \approx \sum_{i=0}^{p} \omega_i \mathbf{g}_{ki}$, such that $\sum_{i=0}^{p} \omega_i = 1$ and $\omega_i > 0, \forall i$;
- updating the model according to $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \mathbf{H}^{-1} \mathbf{g}_k$, where α is a step length obtained from a line search and $\mathbf{H}^{-1}$ is an approximation of the inverse Hessian.

Two major drawbacks are prohibitive computational costs for gradient samples and the quadratic subproblem for weights ω_i, and we circumvolve these issues by implicit approximation of sampling of models in the ball $B_{\epsilon_k}(\mathbf{x}_k)$ and predetermined random weights that satisfy the constraints (positive and sum to one), respectively.

[12] M. Louboutin, F. J. Herrmann, 2017, 79th EAGE Conference & Exhibition, Extending the Search Space of Time-domain Adjoint-state FWI with Randomized Implicit Time Shifts. Date: 2018/12/5 Wen.

12.2.2 Implicit time shift

The gradients of the FWI objective $\Phi_s(\mathbf{m})$ for an acoustic medium:

$$\nabla\Phi_s(\mathbf{m}) = -\sum_{t\in I}\left[\mathrm{diag}(\mathbf{u}[t])(\mathbf{D}^T\mathbf{v}[t])\right]$$

where $\mathbf{m}$: the square slowness; $\mathbf{u}$: the forward wavefield; $\mathbf{v}$: the adjoint wavefield; $\mathbf{D}$: the time derivative discretization matrix; I: the time index set $[1, 2, \ldots, n_t]$.

For a slightly perturbed velocity model $\tilde{\mathbf{m}}$ nearby $\mathbf{m}$,

$$\mathbf{u}(\tilde{\mathbf{m}})[t] \approx \mathbf{u}(\mathbf{m})[t+\tau], \mathbf{v}(\tilde{\mathbf{m}})[t] \approx \mathbf{v}(\mathbf{m})[t-\tau]$$

so that the approximated gradient:

$$\nabla\Phi_s(\tilde{\mathbf{m}}) = -\sum_{t\in I}\left[\mathrm{diag}(\mathbf{u}[t+\tau])(\mathbf{D}^T\mathbf{v}[t-\tau])\right]$$

And by limiting the maximum time shift to $\tau_{max} = \frac{1}{f_0}$, where f_0 is the peak frequency of the source wavelet, guaranty wavefields not to be shifted by more that half a wavelength.

Another way to avoid storage and explicit calculations of gradients is:

$$\overline{\nabla\Phi_s(\mathbf{m})} = -\sum_{t\in\overline{I}}\left[\mathrm{diag}(\overline{\mathbf{u}}[t])(\overline{\mathbf{D}^T\mathbf{v}}[t])\right], \overline{\mathbf{u}} = \sum_{t=t_i}^{t_{i+1}}\sqrt{\alpha_t}\,\mathbf{u}[t], \overline{\mathbf{D}^T\mathbf{v}} = \sum_{t=t_i}^{t_{i+1}}\sqrt{\alpha_t}\,\mathbf{D}^T\mathbf{v}[t]$$

where $\overline{I} = [t_1, t_2, \ldots, t_n]$ are the jitered time sampled from $[1, 2, \ldots, n_t]$, and random weights $\sum\alpha_t = 1$.

13 Schraudolph_2007_AIStat_Stochastic quasi-Newton method[13]

13.1 Introduction

* Accelerate stochastic gradient descent through online adaptation of a gain vector: Schraudolph, 1999 & 2002.
* *************************
* Online implementations of conjugate gradient methods: Møller, 1993; Schraudolph and Graepel, 2003.
* *************************
* Global extened Kalman filtering: Puskorius and Feldkamp, 1991.
* Natural gradient descent: Amari *et al.*, 2000.

Core tools of conventional gradient-based optimization, such as line searches, are not amenable to stochastic approximation.

[13] Nicol N. Schraudolph, Jin Yu, Simon Günter, 2007, 11th International Conference on Artificial Intelligence and Statistics, A Stochastic Quasi-Newton Method for Online Convex Optimization. Date: 2018/12/6 Thu.

13.2 Preliminaries

The objective function $f : \mathbb{R}^n \to \mathbb{R}$:

$$f(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \mathbf{J}\mathbf{J}^T (\boldsymbol{\theta} - \boldsymbol{\theta}^*),$$

where $\boldsymbol{\theta}^* \in \mathbb{R}^n$: the optimal parameter; $\mathbf{J} \in \mathbb{R}^{n \times n}$: the Jacobian matrix. Here the Hessian $\mathbf{H} = \mathbf{J}\mathbf{J}^T$ and the gradient $\nabla f(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$.

A stochastic optimization problem is defined by the data-dependent objective

$$f(\boldsymbol{\theta}, \mathbf{X}) = \frac{1}{2b}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \mathbf{J}\mathbf{X}\mathbf{X}^T \mathbf{J}^T (\boldsymbol{\theta} - \boldsymbol{\theta}^*),$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_b]_{n \times b}$ is a batch of b random input vectors, each drawn i.i.d. (independent identically distribution): $\mathbf{x}_i \sim N(0, b)$, so that $\mathbb{E}[\mathbf{X}\mathbf{X}^T] = b\mathbf{I}$ and

$$\mathbb{E}_{\mathbf{X}}[f(\boldsymbol{\theta}, \mathbf{X})] = \frac{1}{2b}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \mathbf{J}\mathbb{E}[\mathbf{X}\mathbf{X}^T]\mathbf{J}^T (\boldsymbol{\theta} - \boldsymbol{\theta}^*) = f(\boldsymbol{\theta}),$$

and giving rise to the noisy estimates $\mathbf{H} = b^{-1}\mathbf{J}\mathbf{X}\mathbf{X}^T\mathbf{J}^T$ and $\nabla f(\boldsymbol{\theta}, \mathbf{X}) = \mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$. The degree of stochasticity is determined by the batch size b.

As a experiment, we can define an ill-conditioned Jacobian matrix as

$$J_{ij} = \begin{cases} \frac{1}{i+j-1} & \text{if } i \bmod j = 0 \text{ or } j \bmod i = 0, \\ 0 & \text{otherwise.} \end{cases}$$

13.2.1 Stochastic gradient descent (SGD)

Simple stochastic gradient descent:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t),$$

where $\eta_t > 0$ is a scalar gain. The above formula converges to $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$, if provided that

$$\sum_t \eta_t = \infty \text{ and } \sum_t \eta_t^2 < \infty.$$

A commonly used decay schedule:

$$\eta_t = \frac{\tau}{\tau + t}\eta_0,$$

where $\eta_0, \tau > 0$ are tuning parameters.

SGD takes only $O(n)$ space and time per iteration, and suffers from slow convergence on ill-conditioned problems.

13.2.2 Stochastic meta-descent (SMD)

Giving each system parameter its own gain:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \boldsymbol{\eta}_t \cdot \nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t),$$

where the vector gain $\boldsymbol{\eta}_t$ is adapted by

$$\boldsymbol{\eta}_t = \boldsymbol{\eta}_{t-1} \cdot \max\left[\frac{1}{2}, 1 - \mu\nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t) \cdot \boldsymbol{\nu}_t\right],$$

and the auxiliary vector:

$$\boldsymbol{\nu}_{t+1} = \lambda\boldsymbol{\nu}_t - \boldsymbol{\eta}_t \cdot [\nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t) + \lambda\mathbf{H}_t\boldsymbol{\nu}],$$

with another scalar tuning parameter $0 \leqslant \lambda \leqslant 1$.

If $\mathbf{H}_t\boldsymbol{\nu}_t$ can be computed efficiently (Schraudolph, 2002), SMD still takes only $O(n)$ space and time per iteration.

13.2.3 Natural gradient descent (NG)

Incorporate the Riemannian metric tensor $\mathbf{G}_t = \mathbb{E}_{\mathbf{X}}[\nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t)\nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t)^T]$ into the stochastic gradient update:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \hat{\mathbf{G}}_t^{-1} \nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t),$$

with the $\hat{\mathbf{G}}_t$ updated via

$$\hat{\mathbf{G}}_{t+1} = \frac{t-1}{t}\hat{\mathbf{G}}_t + \frac{1}{t}\nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t)\nabla f(\boldsymbol{\theta}_t, \mathbf{X}_t)^T.$$

NG takes $O(n^2)$ space and time per iteration.

13.3 The (L)BFGS algorithm

Here the author puts up more details of algorithms BFGS, oBFGS, LBFGS and oLBFGS in the form of pseudo-codes.

Wave Field Forward

14 ZhangW_2006_GJI_Traction image method[1]

14.1 Introduction

* Use finite difference method (FDM) in rupture dynamics of earthquake source: Madariaga, 1976; Andrews, 1976a & 1976b; Olsen *et al.*, 1997; Madariaga *et al.*, 1998; Cruz-Atienza and Virieux, 2004.
* Use FDM in seismic wave propagation in complex heterogeneous media: Boore, 1972; Kelly *et al.*, 1976; Bayliss *et al.*, 1986; Virieux, 1984 & 1986; Levander, 1988; Graves, 1996; Dai *et al.*, 1995; Zahradnik, 1995.
* Free surface conditions: Jih *et al.*, 1988; Oprsal and Zahradnik, 1999; Ohminato and Chouet, 1997; Robertsson, 1996; Hestholm and Ruud, 1994 & 1998.
* *************************
* Free surface conditions for a planar surface: Gottschammer and Olsen, 2001; Kristek *et al.*, 2002.
* Vacuum method: Boore, 972; Graves, 1996.
* Characteristic variables method: Bayliss *et al.*, 1986.
* Adjusted FD approximations (AFDA) technique: Kristek *et al.*, 2002.
* Stress image method: Levander, 1988; Graves, 1996.
* *************************
* Extend the stress image method with staircase approximation to the general topographic problem in the second-order accurate staggered finite difference scheme: Ohminato and Chouet, 1997.
* Implement the stress image method with staircase approximation to the irregular surface in the fourth-order staggered scheme: Robertsson, 1996; Pitarka and Irikura, 1996.
* *************************
* Vertical grid mapping to match the computational grids with the surface topography in staggered finite difference schemes: Hestholm and Ruud, 1994 & 1998.
* *************************
* Boundary-conforming grid in seismic wave simulation with pseudospectral method: Fornberg, 1988.
* Numerical grid generation: Thompson *et al.*, 1985.
* The original MacCormack scheme with 2nd-order accurate in both time and space: MacCormack, 1969.
* Extend MacCormack scheme to 2nd-order accurate in time and 4th-order accurate in space (2-4 MacCormack scheme): Gottlieb and Turkel, 1976.

[1] Wei Zhang, Xiaofei Chen, 2006, Geophys. J. Int., Traction image method for irregular free surface boundaries in finite difference seismic wave simulation. Date: 2016/10/29 Sat.

* Introduce 2-4 MacCormack scheme into seismic wave modelling: Bayliss *et al.*, 1986 (implement with an operator splitting).
* Use 2-4 MacCormack splitting scheme in seismic wave problems: Xie and Yao, 1988; Tsingas *et al.*, 1990; Vafidis *et al.*, 1992; Dai *et al.*, 1995.
* High-accuracy MacCormack schemes with the DRP/opt MacCormack scheme: Hixon, 1997.
* DRP (dispersion relation preserving) methodology: Tam and Webb, 1993.
* 4-6 LDDRK (low dispersion and dissipation Runge-Kutta) scheme: Hu *et al.*, 1996.
* 4/4 compact MacCormack scheme: Hixon and Turkel, 2000.
* Treat the discontinuous interior interfaces by effective parameters (arithmetic average or harmonic average): Moczo *et al.*, 2002.
* The approximated delta function by Herrmann pseudo-delta functions: Herrmann, 1979; Wang *et al.*, 2001.
* The split-field perfectly matched layer (PML) approach: Bérenger , 1994; Marcinkovich and Olsen, 2003.

14.2 DRP/opt MacCormack scheme

In the DRP scheme, the forward and backward partial difference operators are:

$$\hat{W}_i^F = \frac{1}{\Delta x}\sum_{j=-1}^{3} a_j W_{i+j}$$

$$\hat{W}_i^B = \frac{1}{\Delta x}\sum_{j=-1}^{3} -a_j W_{i-j}$$

where the expansion coefficients are: $a_{-1} = -0.30874, a_0 = -0.6326, a_1 = 1.2330, a_2 = -0.3334, a_3 = 0.04168$ and these coefficients are obtained by minimizing the dissipation error at eight points or more per wavelength.

14.3 Compact MacCormack scheme

The 4/4 compact MacCormack scheme is:

$$\hat{W}_{j-1}^B + 2\hat{W}_j^B = \frac{1}{2\Delta x}(W_{j+1} + 4W_j - 5W_{j-1})$$

$$2\hat{W}_j^F + \hat{W}_{j+1}^F = \frac{1}{2\Delta x}(5W_{j+1} - 4W_j - W_{j-1})$$

where $\hat{W}_j^F$ and $\hat{W}_j^B$ denote the forward and backward difference operators.

14.4 Interior interface conditions

Treat the discontinuous interior interfaces by effective parameters, the density by arithmetic average:

$$\rho_{ij} = \frac{1}{\Delta S} \int_{i-1/2}^{i+1/2} \int_{j-1/2}^{j+1/2} \rho dxdy$$

and the Lamé parameters by harmonic average:

$$\frac{1}{\mu_{ij}} = \frac{1}{\Delta S} \int_{i-1/2}^{i+1/2} \int_{j-1/2}^{j+1/2} \frac{1}{\mu} dxdy$$

$$\frac{1}{\lambda_{ij}} = \frac{1}{\Delta S} \int_{i-1/2}^{i+1/2} \int_{j-1/2}^{j+1/2} \frac{1}{\lambda} dxdy$$

15 ZhangW_2010_Geophy_ADE CFS-PML[2]

15.1 Introduction

* Absorbing boundary conditions (ABC), a proper boundary condition where waves only propagate outward: Clayton and Engquist, 1977; Liao *et al.*, 1984; Bayliss *et al.*, 1986; Higdon, 1986 & 1990; Randall, 1988.
* Absorbing boundary layers (ABL), finite layers to gradually damp wave amplitude: Cerjan *et al.*, 1985 & Sochacki *et al.*, 1987 using the Dirichlet boundary condition.
* Strengths and weaknesses of ABC and ABL: Festa and Vilotte, 2005; Komatitsch and Martin, 2007.
* PML in elastic wave modeling: Chew and Liu, 1996; Hastings *et al.*, 1996; Collino and Tsogka, 2001; Marcinkovich and Olsen, 2003; Wang and Tang, 2003.
* *************************
* Interpret PML: Sacks *et al.*, 1995 & Gedney, 1996 as an artificial anisotropic medium; Chew and Weedon, 1994 & Teixeira and Chew, 2000 as complex coordinate streching.
* Unsplit-field PML implementations: Wang and Tang, 2003 & Komatitsch and Martin, 2007 involving convolution terms; Zeng and Liu, 2004 & Drossaert and Giannopoulos, 2007a involving integral terms; Ramadan, 2003 involving auxiliary differential equations (ADE).
* Modified modal solution to derive PML equations: Hagstrom, 2003 (proposal); Appelö and Kreiss, 2006 (implementation in 2D elastic wave modeling).
* *************************
* Complex-frequency-shifted PML (CFS-PML): Kuzuoglu and Mittra, 1996.
* *************************
* Unsplit-field CFS convolutional-PML (C-PML) involving a convolution term: Roden and Gedney, 2000.

[2] Wei Zhang, Yang Shen, 2010, Geophysics, Unsplit complex frequency-shifted PML implementation using auxiliary differential equations for seismic wave modeling. Date: 2016/11/6 Sun.

* Recursive convolution algorithm: Luebbers and Hunsberger, 1992 (proposal); Komatitsch and Martin, 2007 & Drossaert and Giannopoulos, 2007b (implementation in elastic wave modeling).
* CFS-PML implementation involving integral terms: Drossaert and Giannopoulos, 2007a.
* Recursive integration in C-PML: Giannopoulos, 2008 (1st-order accuracy).
* Trapezoidal rule in recursive integration PML (RIPML): Drossaert and Giannopoulos, 2007a (2nd-order accuracy).
* *************************
* Unsplit-field implementation of the standard PML using auxiliary differential equations (ADE CFS-PML): Ramadan, 2003 (electromagnetic simulation).
* Extend ADE CFS-PML to CFS-PML with 2D alternating-direction-implicit finite difference time domain method: Wang and Liang, 2006.
* *************************
* Adjusted finite difference approximations (AFDA) technique: Kristek *et al.*, 2002 (using a compact finite difference operator and biased finite difference operators).

15.2 Finite difference numerical scheme

For an isotropic elastic medium:

$$\mathbf{v}_{,t} = \frac{1}{\rho}\nabla\cdot\sigma$$

$$\sigma_{,t} = \mathbf{c} : [\nabla\mathbf{v} + (\nabla\mathbf{v})^T]$$

and for the v_x component:

$$\rho v_{x,t} = \sigma_{xx,x} + \sigma_{xy,y} + \sigma_{xz,z}$$

The second-order leapfrog scheme is

$$\sigma^{n+1/2} = \sigma^{n-1/2} + \Delta t\tilde{L}(\mathbf{v}^n)$$

$$\mathbf{v}^{n+1} = \mathbf{v}^n + \Delta t\tilde{L}(\sigma^{n+1/2})$$

15.3 CFS-PML using ADE

Complex stretched coordinate $\tilde{x}$:

$$\tilde{x} = \int_0^x s_x(\eta)d\eta \quad\Rightarrow\quad \frac{\partial\tilde{x}}{\partial x} = s_x(x) \quad\Rightarrow\quad \frac{\partial}{\partial\tilde{x}} = \frac{1}{s_x}\frac{\partial}{\partial x}$$

where s_x is the complex stretching function. As an example in the frequency domain,

$$i\omega\rho\hat{v}_x = \frac{1}{s_x}\frac{\partial\hat{\sigma}_{xx}}{\partial x} + \frac{\partial\hat{\sigma}_{xy}}{\partial y} + \frac{\partial\hat{\sigma}_{xz}}{\partial z}$$

Moreover,

$$s_x(x) = 1 + \frac{d_x(x)}{i\omega} \qquad \text{for the standard PML}$$

$$s_x(x) = \beta_x(x) + \frac{d_x(x)}{\alpha_x(x) + i\omega} \qquad \text{for the CFS-PML}$$

where $d_x \geqslant 0$ is the attenuation factor that reduces exponentially the amplitude, $\alpha_x \geqslant 0$ is the frequency-shifted factor that makes the attenuation frequency-dependent, and $\beta_x \geqslant 1$ is the scaling factor for absorption of evanescent waves and near-grazing incident waves.

The basic idea of ADE implementation of CFS-PML is

$$\frac{1}{s_x} = \frac{1}{\beta_x + \frac{d_x}{\alpha_x + i\omega}} = \frac{\alpha_x + i\omega}{\beta_x(\alpha_x + i\omega) + d_x} = \frac{1}{\beta_x} - \frac{1}{\beta_x}\frac{d_x}{(\alpha_x + i\omega)\beta_x + d_x}$$

Thus,

$$\frac{1}{s_x}\frac{\partial \hat{\sigma}_{xx}}{\partial x} = \frac{1}{\beta_x}\hat{\sigma}_{xx,x} - \frac{1}{\beta_x}\hat{T}^x_{xx}$$

$$\hat{T}^x_{xx} = \frac{d_x}{(\alpha_x + i\omega)\beta_x + d_x}\hat{\sigma}_{xx,x}$$

$$i\omega\hat{T}^x_{xx} + \left(\alpha_x + \frac{d_x}{\beta_x}\right)\hat{T}^x_{xx} = \frac{d_x}{\beta_x}\hat{\sigma}_{xx,x}$$

where T^x_{xx} is the auxiliary memory variable. For the v_x component in the frequency domain,

$$i\omega\rho\hat{v}_x = \frac{1}{\beta_x}\hat{\sigma}_{xx,x} - \frac{1}{\beta_x}\hat{T}^x_{xx} + \hat{\sigma}_{xy,y} + \hat{\sigma}_{xz,z}$$

FT to the time domain, the ADE CFS-PML equation of v_x is:

$$\rho v_{x,t} = \sigma_{xx,x} + \sigma_{xy,y} + \sigma_{xz,z} + \left[\frac{1}{\beta_x} - 1\right]\sigma_{xx,x} - \frac{1}{\beta_x}T^x_{xx}$$

$$T^x_{xx,t} + \left(\alpha_x + \frac{d_x}{\beta_x}\right)T^x_{xx} = \frac{d_x}{\beta_x}\sigma_{xx,x}$$

In the staggered second-order leapfrog scheme, discretize the above equation,

$$\frac{T^{x|n+1}_{xx} - T^{x|n}_{xx}}{\Delta t} + \left(\alpha_x + \frac{d_x}{\beta_x}\right)\frac{T^{x|n+1}_{xx} + T^{x|n}_{xx}}{2} = \frac{d_x}{\beta_x}\delta^{n+1/2}_{xx,x}$$

Then update T^x_{xx} and v_x through

$$T^{x|n+1}_{xx} = \frac{2 - \Delta t\left(\alpha_x + \frac{d_x}{\beta_x}\right)}{2 + \Delta t\left(\alpha_x + \frac{d_x}{\beta_x}\right)}T^{x|n}_{xx} + \frac{\left(\frac{2\Delta t d_x}{\beta_x}\right)}{2 + \Delta t\left(\alpha_x + \frac{d_x}{\beta_x}\right)}\sigma^{n+1/2}_{xx,x}$$

$$T^{x|n+1/2}_{xx} = \frac{T^{x|n+1}_{xx} + T^{x|n}_{xx}}{2} = \frac{2}{2 + \Delta t\left(\alpha_x + \frac{d_x}{\beta_x}\right)}T^{x|n}_{xx} + \frac{\left(\frac{\Delta t d_x}{\beta_x}\right)}{2 + \Delta t\left(\alpha_x + \frac{d_x}{\beta_x}\right)}\sigma^{n+1/2}_{xx,x}$$

$$v^{n+1}_x = v^n_x + \frac{\Delta t}{\rho}(\sigma^{n+1/2}_{xx,x} + \sigma^{n+1/2}_{xy,y} + \sigma^{n+1/2}_{xz,z}) + \frac{\Delta t}{\rho}\left(\frac{1}{\beta_x} - 1\right)\sigma^{n+1/2}_{xx,x} - \frac{\Delta t}{\rho\beta_x}T^{x|n+1/2}_{xx}$$

15.4 Free-surface boundary conditions

At the flat surface, the free-surface boundary condition requires

$$\sigma_{zz} = 0, \quad \sigma_{yz} = 0, \quad \sigma_{xz} = 0$$

$$v_{z,z} = -\frac{\lambda}{\lambda + 2\mu}v_{x,x} - \frac{\lambda}{\lambda + 2\mu}v_{y,y}$$

Taking into the stress-strain relation, update σ_{xx} and σ_{yy} at the free surface through

$$\sigma_{xx,t} = (\lambda + 2\mu)v_{x,x} + \lambda v_{y,y} + \lambda\Big[-\frac{\lambda}{\lambda+2\mu}v_{x,x} - \frac{\lambda}{\lambda+2\mu}v_{y,y}\Big]$$

$$\sigma_{yy,t} = \lambda v_{x,x} + (\lambda + 2\mu)v_{y,y} + \lambda\Big[-\frac{\lambda}{\lambda+2\mu}v_{x,x} - \frac{\lambda}{\lambda+2\mu}v_{y,y}\Big]$$

Under the intersection of the free surface and the PML, (different form with eq.A-13 and A-14 in the original paper)

$$\begin{aligned}
\sigma_{xx,t} =& (\lambda + 2\mu)v_{x,x} + \lambda v_{y,y} {\color{lightgray}+\lambda v_{z,z}} + (\lambda+2\mu)\Big[\frac{1}{\beta_x}-1\Big]v_{x,x} + \lambda\Big[\frac{1}{\beta_y}-1\Big]v_{y,y} - (\lambda+2\mu)\frac{1}{\beta_x}V_x^x - \lambda\frac{1}{\beta_y}V_y^y \\
&-\frac{\lambda^2}{\lambda+2\mu}\Big\{{\color{lightgray}\Big[}\frac{1}{\beta_x}{\color{lightgray}-1\Big]}v_{x,x} + {\color{lightgray}\Big[}\frac{1}{\beta_y}{\color{lightgray}-1\Big]}v_{y,y} - \frac{1}{\beta_x}V_x^x - \frac{1}{\beta_y}V_y^y\Big\} \\
\sigma_{yy,t} =& \lambda v_{x,x} + (\lambda + 2\mu)v_{y,y} {\color{lightgray}+\lambda v_{z,z}} + \lambda\Big[\frac{1}{\beta_x}-1\Big]v_{x,x} + (\lambda+2\mu)\Big[\frac{1}{\beta_y}-1\Big]v_{y,y} - \lambda\frac{1}{\beta_x}V_x^x - (\lambda+2\mu)\frac{1}{\beta_y}V_y^y \\
&-\frac{\lambda^2}{\lambda+2\mu}\Big\{{\color{lightgray}\Big[}\frac{1}{\beta_x}{\color{lightgray}-1\Big]}v_{x,x} + {\color{lightgray}\Big[}\frac{1}{\beta_y}{\color{lightgray}-1\Big]}v_{y,y} - \frac{1}{\beta_x}V_x^x - \frac{1}{\beta_y}V_y^y\Big\}
\end{aligned}$$

15.5 Optimal parameters

The d usually is zero at the PML-interior interface and maximum at the exterior boundary, β is one at the PML-interior interface and maximum at the exterior boundary, and α is maximum at the PML-interior interface and gradually reduces to zero at the exterior boundary.

The commonly used optimal parameters is p-order polynomial scaling functions:

$$\alpha_x = \alpha_0\Big[1 - \Big(\frac{x}{L}\Big)^{p_\alpha}\Big]$$

$$d_x = d_0\Big(\frac{x}{L}\Big)^{p_d}$$

$$\beta_x = 1 + (\beta_0 - 1)\Big(\frac{x}{L}\Big)^{p_\beta}$$

where x is the distance to the PML-interior interface and L is the width of the PML layer. The parameters p_α, p_d and p_β typically range from $2 \sim 4$, and 2 is commonly used, e.g. $p_d = 2$, $p_\beta = 2$, and $p_\alpha = 1$ (the linear variation of α for $p_\alpha = 1$ gets a more pronounced decay of energy).

The α_0 is recommended to be πf_c (Festa and Vilotte, 2005), where f_c is the dominant frequency of the source time function.

The d_0 is (Collino and Tsogka, 2001):

$$d_0 = -\frac{(p_d + 1)c_p}{2L}\ln R$$

where c_p is the compressional wave speed and R is the theoretical reflection coefficient for a normal-incident plane P-wave with a Dirichlet condition ($\mathbf{v} = 0$ and $\sigma = 0$) at the exterior boundary of the PML layer. R for an N cell size PML layer is:

$$\log_{10}(R) = -\frac{\log_{10}(N) - 1}{\log_{10}(2)} - 3$$

For oblique incident waves, a larger d_0 is needed to obtain optimal damping.

The optimal β_0 is

$$\beta_0 = \frac{C}{0.5 \cdot \mathrm{PPW}_0 \cdot \Delta h f_c}$$

where C is wave velocity, PPW_0 is the minimal PPW requirement of the numerical scheme, Δh is grid spacing, and f_c is source dominant frequency.

15.6 Complete ADE CFS-PML equations

The complete ADE CFS-PML equations for the velocity-stress equations are:

$$
\begin{cases}
\sigma_{xx,t} =(\lambda+2\mu)v_{x,x}+\lambda v_{y,y}+\lambda v_{z,z}+(\lambda+2\mu)\Big[\dfrac{1}{\beta_x}-1\Big]v_{x,x}+\lambda\Big[\dfrac{1}{\beta_y}-1\Big]v_{y,y}+\lambda\Big[\dfrac{1}{\beta_z}-1\Big]v_{z,z}\\
\qquad -(\lambda+2\mu)\dfrac{1}{\beta_x}V_x^x-\lambda\dfrac{1}{\beta_y}V_y^y-\lambda\dfrac{1}{\beta_x}V_z^z\\
\sigma_{yy,t} =\lambda v_{x,x}+(\lambda+2\mu)v_{y,y}+\lambda v_{z,z}+\lambda\Big[\dfrac{1}{\beta_x}-1\Big]v_{x,x}+(\lambda+2\mu)\Big[\dfrac{1}{\beta_y}-1\Big]v_{y,y}+\lambda\Big[\dfrac{1}{\beta_z}-1\Big]v_{z,z}\\
\qquad -\lambda\dfrac{1}{\beta_x}V_x^x-(\lambda+2\mu)\dfrac{1}{\beta_y}V_y^y-\lambda\dfrac{1}{\beta_z}V_z^z\\
\sigma_{zz,t} =\lambda v_{x,x}+\lambda v_{y,y}+(\lambda+2\mu)v_{z,z}+\lambda\Big[\dfrac{1}{\beta_x}-1\Big]v_{x,x}+\lambda\Big[\dfrac{1}{\beta_y}-1\Big]v_{y,y}+(\lambda+2\mu)\Big[\dfrac{1}{\beta_z}-1\Big]v_{z,z}\\
\qquad -\lambda\dfrac{1}{\beta_x}V_x^x-\lambda\dfrac{1}{\beta_y}V_y^y-(\lambda+2\mu)\dfrac{1}{\beta_z}V_z^z\\
\sigma_{xy,t} =\mu(v_{x,y}+v_{y,x})+\mu\Big(\Big[\dfrac{1}{\beta_y}-1\Big]v_{x,y}+\Big[\dfrac{1}{\beta_x}-1\Big]v_{y,x}\Big)-\mu\Big(\dfrac{1}{\beta_y}V_x^y+\dfrac{1}{\beta_x}V_y^x\Big)\\
\sigma_{xz,t} =\mu(v_{x,z}+v_{z,x})+\mu\Big(\Big[\dfrac{1}{\beta_z}-1\Big]v_{x,z}+\Big[\dfrac{1}{\beta_x}-1\Big]v_{z,x}\Big)-\mu\Big(\dfrac{1}{\beta_z}V_x^z+\dfrac{1}{\beta_x}V_z^x\Big)\\
\sigma_{yz,t} =\mu(v_{y,z}+v_{z,y})+\mu\Big(\Big[\dfrac{1}{\beta_z}-1\Big]v_{y,z}+\Big[\dfrac{1}{\beta_y}-1\Big]v_{z,y}\Big)-\mu\Big(\dfrac{1}{\beta_z}V_y^z+\dfrac{1}{\beta_y}V_z^y\Big)\\
\rho v_{x,t} =\sigma_{xx,x}+\sigma_{xy,y}+\sigma_{xz,z}+\Big[\dfrac{1}{\beta_x}-1\Big]\sigma_{xx,x}+\Big[\dfrac{1}{\beta_y}-1\Big]\sigma_{xy,y}+\Big[\dfrac{1}{\beta_z}-1\Big]\sigma_{xz,z}\\
\qquad -\dfrac{1}{\beta_x}T_{xx}^x-\dfrac{1}{\beta_y}T_{xy}^y-\dfrac{1}{\beta_z}T_{xz}^z\\
\rho v_{y,t} =\sigma_{xy,x}+\sigma_{yy,y}+\sigma_{yz,z}+\Big[\dfrac{1}{\beta_x}-1\Big]\sigma_{xy,x}+\Big[\dfrac{1}{\beta_y}-1\Big]\sigma_{yy,y}+\Big[\dfrac{1}{\beta_z}-1\Big]\sigma_{yz,z}\\
\qquad -\dfrac{1}{\beta_x}T_{xy}^x-\dfrac{1}{\beta_y}T_{yy}^y-\dfrac{1}{\beta_z}T_{yz}^z\\
\rho v_{z,t} =\sigma_{xz,x}+\sigma_{yz,y}+\sigma_{zz,z}+\Big[\dfrac{1}{\beta_x}-1\Big]\sigma_{xz,x}+\Big[\dfrac{1}{\beta_y}-1\Big]\sigma_{yz,y}+\Big[\dfrac{1}{\beta_z}-1\Big]\sigma_{zz,z}\\
\qquad -\dfrac{1}{\beta_x}T_{xz}^x-\dfrac{1}{\beta_y}T_{yz}^y-\dfrac{1}{\beta_z}T_{zz}^z
\end{cases}
$$

where the auxiliary differential equations for the memory variables damping along x, y and z are:

$$
x\begin{cases}
V_{x,t}^x+\Big(\alpha_x+\dfrac{d_x}{\beta_x}\Big)V_x^x=\dfrac{d_x}{\beta_x}v_{x,x}, \quad V_{y,t}^x+\Big(\alpha_x+\dfrac{d_x}{\beta_x}\Big)V_y^x=\dfrac{d_x}{\beta_x}v_{y,x}, \quad V_{z,t}^x+\Big(\alpha_x+\dfrac{d_x}{\beta_x}\Big)V_z^x=\dfrac{d_x}{\beta_x}v_{z,x}\\
T_{xx,t}^x+\Big(\alpha_x+\dfrac{d_x}{\beta_x}\Big)T_{xx}^x=\dfrac{d_x}{\beta_x}\sigma_{xx,x}, \quad T_{xy,t}^x+\Big(\alpha_x+\dfrac{d_x}{\beta_x}\Big)T_{xy}^x=\dfrac{d_x}{\beta_x}\sigma_{xy,x}, \quad T_{xz,t}^x+\Big(\alpha_x+\dfrac{d_x}{\beta_x}\Big)T_{xz}^x=\dfrac{d_x}{\beta_x}\sigma_{xz,x}
\end{cases}
$$

$$
y\begin{cases}
V_{x,t}^y+\Big(\alpha_y+\dfrac{d_y}{\beta_y}\Big)V_x^y=\dfrac{d_y}{\beta_y}v_{x,y}, \quad V_{y,t}^y+\Big(\alpha_y+\dfrac{d_y}{\beta_y}\Big)V_y^y=\dfrac{d_y}{\beta_y}v_{y,y}, \quad V_{z,t}^y+\Big(\alpha_y+\dfrac{d_y}{\beta_y}\Big)V_z^y=\dfrac{d_y}{\beta_y}v_{z,y}\\
T_{xy,t}^y+\Big(\alpha_y+\dfrac{d_y}{\beta_y}\Big)T_{xy}^y=\dfrac{d_y}{\beta_y}\sigma_{xy,y}, \quad T_{yy,t}^y+\Big(\alpha_y+\dfrac{d_y}{\beta_y}\Big)T_{yy}^y=\dfrac{d_y}{\beta_y}\sigma_{yy,x}, \quad T_{yz,t}^y+\Big(\alpha_y+\dfrac{d_y}{\beta_y}\Big)T_{yz}^y=\dfrac{d_y}{\beta_y}\sigma_{yz,y}
\end{cases}
$$

$$
z\begin{cases}
V_{x,t}^z+\Big(\alpha_z+\dfrac{d_z}{\beta_z}\Big)V_x^z=\dfrac{d_z}{\beta_z}v_{x,z}, \quad V_{y,t}^z+\Big(\alpha_z+\dfrac{d_z}{\beta_z}\Big)V_y^z=\dfrac{d_z}{\beta_z}v_{y,z}, \quad V_{z,t}^z+\Big(\alpha_z+\dfrac{d_z}{\beta_z}\Big)V_z^z=\dfrac{d_z}{\beta_z}v_{z,z}\\
T_{xz,t}^z+\Big(\alpha_z+\dfrac{d_z}{\beta_z}\Big)T_{xz}^z=\dfrac{d_z}{\beta_z}\sigma_{xz,z}, \quad T_{yz,t}^z+\Big(\alpha_z+\dfrac{d_z}{\beta_z}\Big)T_{yz}^z=\dfrac{d_z}{\beta_z}\sigma_{yz,z}, \quad T_{zz,t}^z+\Big(\alpha_z+\dfrac{d_z}{\beta_z}\Big)T_{zz}^z=\dfrac{d_z}{\beta_z}\sigma_{zz,z}
\end{cases}
$$

Index

Concept

Kernel

banana-doughnut kernel, 20, 34, 38
Born kernel, see *banana-doughnut kernel*
event kernel, 38
finite-frequency kernel, see *banana-doughnut kernel*
misfit kernel, 38
sensitivity kernel, see *banana-doughnut kernel*
traveltime misfit kernel, 21
waveform misfit kernel, 17

Seismic

membrane wave, 35
mineralogical seismic anisotropy, 14
structural seismic anisotropy, 14

Inversion

Iteration

full Newton method, 7
Gauss-Newton method, 7, 35
Gradient method, 5
gradient method, 33
integrated stochastic gradient descent, 46
natural gradient descent, 49
Newton method, 6, 33
quasi-Newton algorithm, 12, see *Gauss-Newton method*
stochastic gradient descent, 45, 48
stochastic LBFGS, 45
stochastic meta-descent, 48
stochastic oLBFGS, 46

Other

Algorithm

conjugate gradient, 37
gradient sampling algorithm, 46
LBFGS, 45

Conclusion

benefit and disadvantage of the adjoint approach, 28
different results from different data sets, 27
four criteria influenced selection of earthquakes, 28

Discussion

advantage and drawback of different measurements, 44
gradient image .vs. imgration, 11

Method

automated phase-picking, 39
cup-and-paste, 26
Hilbert transform, 40
Lagrange multiplier, 29

Model

AK135: 1-D, 13
crust2.0: 3-D crust, 13
S20RTS: 3-D mantle, 39

Software

FLEXWIN, 26