Full Waveform Inversion in Theory

Romain Brossier and Jean Virieux
romain.brossier@obs.ujf-grenoble.fr jean.virieux@obs.ujf-grenoble.fr

ISTerre, Université Joseph Fourier, Grenoble, France
SEISCOPE Consortium

12^{th} International Congress of the Brazilian Geophysical Society
Short Course on Full Waveform Inversion
15 August 2011

Outline

Some references

This part of the course is mainly based on the following paper and presentation

- Virieux J. and S. Operto, [2009] An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6), WCC1-WCC26.
- S. Operto [2006] Seismic imaging by frequency-domain full-waveform inversion Part 1 : synthetic examples. SEISCOPE SUMMER SCHOOL

Outline

1. Introduction
2. Local search
3. Regularization
4. Conclusion

Inverse problem ... various terminologies

- What is seismic tomography ?
- What is seismic inversion ?

- Seismic tomography ... reflection tomography and, therefore, time
- Seismic inversion ... MVA/FWI and, therefore, amplitude

- tomography ; a generic term : model characterization from various observations
- inversion ; a technical term : given a transformation between model and observations, provide the inverse transformation from observations to models

- linear formulation
- non-linear formulation
- linearized formulation

Linear inversion problem

- Linear forward problem $d = Gm$: operator G does not depend on m
- Example : Reflection travel time

$$t_i^2 = t_0^2 + x_i^2 / v_{nmo}^2$$

with respect to two parameters $(t_0^2, 1/v_{NMO}^2)$ (X^2-T^2 method)

- Sensitivity matrix G, also known as Fréchet derivative or Jacobian matrix
  G has two columns $G_{i1} = 1$ and $G_{i2} = x_i^2$

- Linear algebra machinery Ax=b ... to be discussed later on ...

Non-linear inversion problem

- Non-linear forward problem $d = \mathcal{G}(m)$: operator $\mathcal{G}$ does depend on m
- Example : Acoustic wave equation

$$\frac{\omega^2}{\kappa(x,z)}P(x,z,\omega)+\frac{\partial}{\partial x}\left(\frac{1}{\rho(x,z)}\frac{\partial P(x,z,\omega)}{\partial x}\right)+\frac{\partial}{\partial z}\left(\frac{1}{\rho(x,z)}\frac{\partial P(x,z,\omega)}{\partial z}\right) = S(x,z,\omega)$$

- Operator $\mathcal{G}(\kappa(x,z), \rho(x,z))$ is a non-linear operator
- Only the forward problem is required in this truely non-linear approach
- Need of an efficient forward problem ...
- Finding model parameters ?
  - Global search : *exhaustive* sampling (grid search, Monte-Carlo)

(Backus and Gilbert, 1967)

Non-linear inversion problem

- Non-linear forward problem $d = \mathcal{G}(m)$: operator $\mathcal{G}$ does depend on m
- Example : Acoustic wave equation

$$\frac{\omega^2}{\kappa(x,z)} P(x,z,\omega) + \frac{\partial}{\partial x}\left(\frac{1}{\rho(x,z)} \frac{\partial P(x,z,\omega)}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho(x,z)} \frac{\partial P(x,z,\omega)}{\partial z}\right) = S(x,z,\omega)$$

- Operator $\mathcal{G}(\kappa(x,z), \rho(x,z))$ is a non-linear operator
- Only the forward problem is required in this truely non-linear approach
- Need of an efficient forward problem ...
- Finding model parameters ?
  - Global search : *exhaustive* sampling (grid search, Monte-Carlo)
  - Semi-global search : Simulated annealing, Genetic algorithm, Ant search, Particle swarm optimization

(Backus and Gilbert, 1967)

Non-linear inversion problem

- Non-linear forward problem $d = \mathcal{G}(m)$: operator $\mathcal{G}$ does depend on m
- Example : Acoustic wave equation

$$\frac{\omega^2}{\kappa(x,z)} P(x,z,\omega) + \frac{\partial}{\partial x}\left(\frac{1}{\rho(x,z)} \frac{\partial P(x,z,\omega)}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho(x,z)} \frac{\partial P(x,z,\omega)}{\partial z}\right) = S(x,z,\omega)$$

- Operator $\mathcal{G}(\kappa(x,z), \rho(x,z))$ is a non-linear operator
- Only the forward problem is required in this truely non-linear approach
- Need of an efficient forward problem ...
- Finding model parameters ?
  - Global search : *exhaustive* sampling (grid search, Monte-Carlo)
  - Semi-global search : Simulated annealing, Genetic algorithm, Ant search, Particle swarm optimization
  - Local search : Simplex method

(Backus and Gilbert, 1967)

Linearized inverse problem

- Selection of a model m_0
- Differentiability : derivative $\frac{\partial \mathcal{G}}{\partial m}(m_0)$ exists

Linearized search is based on a local linearized forward problem

$$
\begin{gathered}
d = \mathcal{G}(m_0) + \partial \mathcal{G}/\partial m_0 (m - m_0) \\
d_0 = \mathcal{G}(m_0) \\
d - d_0 = \partial \mathcal{G}/\partial m_0 (m - m_0) \\
\delta d = \partial \mathcal{G}/\partial m_0 \delta m
\end{gathered}
$$

Forward problem $\mathcal{G}(m_0)$ and first-derivative (gradient) $\frac{\partial \mathcal{G}}{\partial m}(m_0)$ estimated at m_0 as non-linear relations

Model m_0 fixed (linear) or changing (linearized) during iterative inversion

- Back to linear algebra machinery
- Except that the forward problem is non-linear

Misfit function

Objective function or misfit function

$$
\begin{aligned}
\ell^2 norm: \quad \mathcal{C}^2(m) &= \frac{1}{2}\left|\left|d^{obs} - d^{syn}(m)\right|\right|^2 \\
\ell^1 norm: \quad \mathcal{C}^1(m) &= \frac{1}{2}\left|d^{obs} - d^{syn}(m)\right| \\
\ell^p norm: \quad \mathcal{C}^p(m) &= \frac{1}{2}\sqrt[p]{\left|\left|d^{obs} - d^{syn}(m)\right|\right|^p}
\end{aligned}
$$

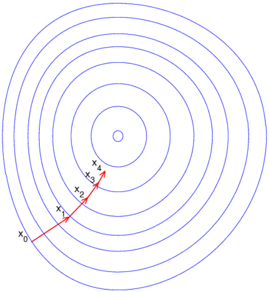
When considering complex numbers, we define
$\mathcal{C}^2(m) = \frac{1}{2}(d^{obs} - d^{syn}(m))^\dagger(d^{obs} - d^{syn}(m))$
or
$\mathcal{C}^2(m) = \frac{1}{2}(d^{obs} - d^{syn}(m))^t(d^{obs} - d^{syn}(m))^*$
$\dagger$ adjoint $=$ transpose and conjugate
^{t} transpose
$*$ conjugate
Correct normalization by the number of data is required in these formulae

Outline

Best fitting model x in the vicinity of a running model x_i

$x_{i+1} = x_i + \delta x_i$

δx_i is the model perturbation



Locally around the model m^i

$$\begin{aligned}\mathcal{C}(m) = \mathcal{C}(m_0 + \Delta m) = \mathcal{C}(m^i) \quad &+ \quad \frac{\partial \mathcal{C}}{\partial m_j}(m^i)\delta m_j \\ &+ \quad \frac{1}{2}\frac{\partial^2 \mathcal{C}}{\partial m_j \partial m_k}(m^i)\delta m_j \delta m_k + \mathcal{O}(m^3)\end{aligned}$$

in explicit form

$$\frac{\partial \mathcal{C}}{\partial m_j}(m) \sim \frac{\partial \mathcal{C}}{\partial m_j}(m^i) + \frac{\partial^2 \mathcal{C}}{\partial m_j \partial m_k}(m^i)\delta m_k$$

in matrix form

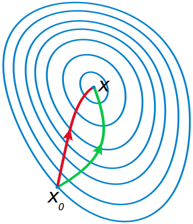
$$\frac{\partial C}{\partial m}(m) \sim \frac{\partial C}{\partial m}(m^i) + \frac{\partial^2 \mathcal{C}}{\partial m^2}(m^i)\delta m$$

Local search (3)

$\frac{\partial C}{\partial m}(m) = 0$ around the model m^i gives

$$\delta m = -\left[\frac{\partial^2 C}{\partial m^2}(m^i)\right]^{-1} \frac{\partial C}{\partial m}(m^i)$$

$$\nabla \mathcal{C}^i = \frac{\partial C}{\partial m}(m^i)$$
$$\mathcal{H}^i = \frac{\partial^2 C}{\partial m^2}(m^i)$$



Line green is gradient descent while line red is the Newton method which uses the curvature for a more direct road

Gradient $\nabla \mathcal{C}^i$ defines the local direction of the steepest ascent of the point m^i

Hessian $\mathcal{H}^i$ defines the local curvature of the misfit function

- ☐ linear problem : the relation is exact
- ☐ Dimension analysis OK when the Hessian is included ...

$$\nabla C = \sum_{s=1}^{N_s} \nabla C^s \quad \text{sum over sources}$$

$$\nabla C_j^s = \frac{1}{2} \sum_{n=1}^{N_d} \frac{\partial (d_n^{obs} - u_n(m^i))^* (d_n^{obs} - u_n(m)}{\partial m_j}$$

$$\nabla C_j^s = -\frac{1}{2} \sum_{n=1}^{N_d} (d_n^{obs} - u_n(m^i))^* \frac{\partial u_n(m^i)}{\partial m_j} + \frac{\partial u_n(m^i)^*}{\partial m_j} (d_n^{obs} - u_n(m^i))$$

Because $x + \bar{x} = 2\mathcal{R}(x)$

$$\nabla C_j^s = -\sum_{n=1}^{N_d} \mathcal{R} \left[\frac{\partial u_n(m^i)}{\partial m_j} (d_n^{obs} - u_n(m^i))^* \right]$$

$$\text{in vector form} \quad \nabla C^s = -\mathcal{R} \left[\frac{\partial u(m^i)^t}{\partial m} \cdot (d^{obs} - u(m^i))^* \right]$$

Gradient estimation for the ℓ^1 norm

Because $\mathcal{C}_{\ell^1} \sim \sqrt{\mathcal{C}_{\ell^2}}$, we have

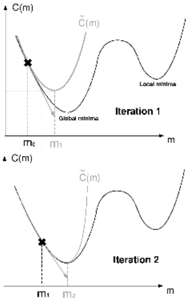
$$\frac{\partial C_{\ell^1}}{\partial m_l} = -\frac{1}{2}\frac{1}{C_{\ell^1}}\frac{\partial C_{\ell^2}}{\partial m_l}$$

$$\nabla C^s_{\ell^1} = \frac{\sqrt{2}}{4}\mathcal{R}\left[\frac{\partial u(m^i)^t}{\partial m} \cdot \frac{(d^{obs} - u(m^i)^*}{|d^{obs} - u(m^i)|}\right]$$

One can see the effect of residues $d^{obs} - u(m^i)$ through a normalized amplitude.

Sensitivity or Jacobian matrix - Fréchet derivative

$$\nabla C(m^i) = -\sum_{s=1}^{N_s} \mathcal{R}\left[J_s^t \Delta d_s^*\right] \quad (1)$$



where $\Delta d_s = d^{obs} - u(m^i) = d^{obs} - Rv(m^i)$ with R is the restriction projector to receivers of the field v computed through the wave equation $Av = b$

where the dimension of the jacobian matrix is $dim(J) = [N_s N_d \times M]$.

Hessian expression

☐ Newton approach

$$\frac{\partial^2 C}{\partial m_k \partial m_l}(m^i) = -\sum_{n=1}^{N_d} \frac{\partial u_n}{\partial m_k}(m^i)^* \frac{\partial u_n}{\partial m_l}(m^i) + \frac{\partial^2 u_n}{\partial m_k \partial m_l}(m^i)(d_n^{obs} - u_n(m^i))^*$$

$$\frac{\partial^2 C}{\partial m^2}(m^i) = -\mathcal{R}(J^t(m^i)J(m^i)^*) - \mathcal{R}\left[\sum_{k=1}^{M} \frac{\partial J^t}{\partial m_k} \Delta d^*\right]$$

$$\frac{\partial^2 C}{\partial m^2}(m^i) = -\mathcal{R}(J^t(m^i)J(m^i)^*) - \mathcal{R}\left[\frac{\partial J^t}{\partial m} \cdot (\Delta d_1 ... \Delta d_{N_s})^*\right]$$

Multi-scattering is considered through the influence of m_k over m_l

Check the dimensions of the matrix $dim(J^t) = [M \times N_s N_d]$

☐ Gauss-Newton approach

$$\frac{\partial^2 C}{\partial m^2}(m^i) \sim -\mathcal{R}\left(J^t(m^i)J(m^i)^*\right)$$

The normal system

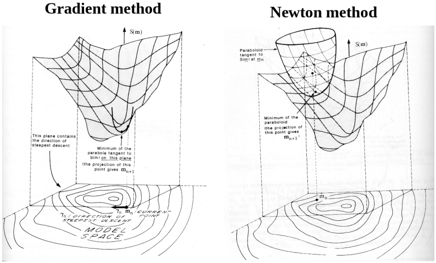
☐ Newton approach

$$\delta m^{i+1} = -\left(\mathcal{R}\left[J^t(m^i)J^*(m^i) + \frac{\partial J^t}{\partial m}(m^i)\cdot(\Delta d_1 ... \Delta d_{N_s})^*\right]\right)^{-1}\mathcal{R}\left[J^t(m^i)\Delta d^*(m^i)\right]$$

☐ Gauss-Newton approach

$$\delta m^{i+1} \sim -\mathcal{R}\left[J^t(m^i)J^*(m^i)\right]^{-1}\mathcal{R}\left[J^t(m^i)\Delta d^*(m^i)\right]$$

Gradient method versus Newton method



From Tarantola (1987)

$$\delta m_i^{Gradient} = -\alpha \mathcal{R}(J_i^t \Delta d_i^*)$$
$$\delta m_i^{Gauss} = -\mathcal{R}\left[J_i^t J_i\right]^{-1} \mathcal{R}(J_i^t \Delta d_i^*)$$
$$\delta m_i^{Newton} = -\left[\mathcal{H}_i\right]^{-1} \mathcal{R}(J_i^t \Delta d_i^*)$$

where the iteration is taken as an index
Gradient method : α scaling factor distance and proper units
Approximate Gauss-Newton method : diagonal Hessian matrix
Quasi-Newton method : better estimation of the Hessian or its inverse or its effect onto the gradient (DFP ; BFGS ; L-BFGS ; L-BFGS-B ; KKT ...) (Pratt et al., 1998; Nocedal and Wright, 1999)

Adjoint approach for the gradient estimation

No need to estimate the Jacobian matrix J for the gradient estimation. At a given iteration i, we have

$$\nabla C = \mathcal{R}\left[J^t \Delta d^*\right].$$

The adjoint formulation allows the writing for each source

$$\nabla C_l^s = \mathcal{R}\left[u^t (\frac{\partial B}{\partial m_l})^t r_b^*\right] \quad \text{with} \quad Bu = s$$

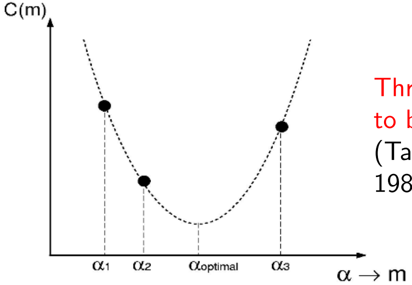
- $[M \times 1]$ u^s incident wave for the source s : $B^{-1}s$
- $[M \times 1]$ r_b^s backpropagated residues : $B^{-1}\Delta d$
- $[M \times M]$ $\frac{\partial B}{\partial m_l}$ sparse diffraction kernel

$$\nabla C = \sum_{s=1}^{N_s} \mathcal{R}\left[{u^s}^t (\frac{\partial B}{\partial m})^t {r_b^s}^*\right],$$

Two forward problems to be solved at least (Plessix, 2006)

$$\delta m_i^{gradient} = -\alpha \mathcal{R}\left[J_i^t \Delta d_i^*\right]$$

step length α parabolic fitting



Three forward problems
to be solved at least
(Tarantola, 1984; Gauthier et al., 1986)

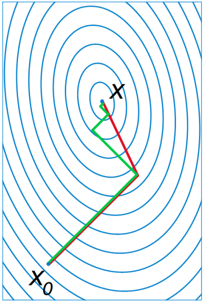
Improved method : approximate Gauss-Newton method but need estimation of diagonal terms of Jacobian matrix

$$\delta m_i^{gradient} = -Diag(\mathcal{R}\left[J_i^t J_i^*\right])^{-1}\mathcal{R}\left[J_i^t \Delta d_i^*\right]$$

N_d forward problems for each source to be solved at most

Conjugate Gradient method

$$p^i = \nabla C^i + \beta^i p^{i-1} \quad (2)$$



p^i is the new search direction from

- previous direction p^{i-1}
- new gradient ∇C^i
- factor
  $\beta^i =< \nabla C^i - \nabla C^{i-1} | \nabla C^i > / (\nabla C^i)^2$

(Polak and Ribière, 1969; Mora, 1987)

green line is the gradient descent while red is the conjugate gradient

Quasi-Newton methods

$$\delta m_i \sim \mathcal{H}_i^{-1} \mathcal{R}(J^t \Delta d^*)_i$$

with notation $\nabla C(m_i + \delta m_i) \sim \nabla C(m_i) + \mathcal{H}_i \delta m_i$, we have

☐ Davidon-Fletcher-Powel (DFP) formulae (secant formulae)

$$\begin{aligned} \mathcal{H}_{i+1} &= (I - \gamma_i y_i \delta m_i^t) \mathcal{H}_i (I - \gamma_i \delta m_i^t y_i) + \gamma_i y_i y_i^t \\ \mathcal{H}_{i+1}^{-1} &= H_i^{-1} + \gamma_i \delta m_i \delta m_i^t - \frac{\mathcal{H}_i y_i (\mathcal{H}_i y_i)^t}{y_i^t \mathcal{H}_i y_i} \end{aligned}$$

with $y_i = \nabla C(m_i + \delta m_i) - \nabla C(m_i)$ and $\gamma_i = 1/y_i^t \delta m_i$

while the more efficient BFGS formula has a very analog expression

☐ Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula

$$\begin{aligned} \mathcal{H}_{i+1} &= \mathcal{H}_i + \gamma_i y_i y_i^t - \frac{\mathcal{H}_i \delta m_i (\mathcal{H}_i \delta m_i)^t}{\delta m_i^t \mathcal{H}_i \delta m_i} \\ \mathcal{H}_{i+1}^{-1} &= (I - \gamma_i y_i \delta m_i^t) \mathcal{H}_i^{-1} (I - \gamma_i \delta m_i^t y_i) + \gamma_i \delta m_i \delta m_i^t \end{aligned}$$

Quasi-Newton methods cont'd

☐ Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) formula
This approach is different on how matrix-vector multiplication is performed for finding the search direction

☐ Limited-memory Broyden–Fletcher–Goldfarb–Shanno for bound constrained optimization (L-BFGS-B) formula
when bound limites for model parameters are required (Nocedal, 1980).

☐ Karush-Kuhn-Tucker conditions (KKT) formula
with inegality and equality constraints ...
Other conditions for allowing various constraints (the use of Lagrange multipliers)

Outline

Ill-posed inversion problem

- Seismic data : incomplete illumination of the medium
- under-determined problem
- over-determined problem
- mixed-determined problem

- Remedy : regularization (Tikhonov and Arsenin, 1977)
- data space W_d weighting operator (quality and/or compression)
- model space W_m weighting operator (could be a combination)

Weighted least-square problem ℓ^2 ℓ^2

$$\mathcal{C} = \mathcal{C}_{data} + \mathcal{C}_{model} = \frac{1}{2}(\Delta d^{\dagger} W_d \Delta d) + \frac{1}{2}\epsilon(\Delta m^{\dagger} W_m \Delta m)$$

W_d : data are often not correlated - diagonal matrix
W_m : model parameter weighting smoothing or roughing or both ...
ϵ : added this extra parameter for easiness (although could be included in W_m)
sometimes called Ridge regression or Tikhonov regularisation ($W_m \sim I$)
related to the Levenberg-Marquardt algorithm

$[dim(W_d)]=$ inverse of the square of the data units
$[dim[W_m)]=$ inverse of the square of the model units

$\Delta m = m^i - m_{prior}$ where m_{prior} might take various values
as m^i or m^{i-1} or m_{ref}.
Higher-order Tikhonov (W_m is a first-order or second-order differential operator)

Least Absolute Shrinkage and Selection Operator "LASSO" L2L1 regression ℓ^2 ℓ^1

$$\mathcal{C} = \mathcal{C}_{data} + \mathcal{C}_{model} = \frac{1}{2}(\Delta d^\dagger W_d \Delta d) + \frac{1}{2}\epsilon|\sqrt{W_m}\Delta m|$$

W_d : Data are often not correlated - diagonal matrix
$W_m = L$: In the Total Variation regression, the operator L is the first-order differential operator
For TV regression $\nabla C_{model} \sim -\nabla(\frac{\nabla m}{|\nabla m| + \beta^2})$
β a smoothing factor in order to avoid singularities in the model gradient estimation
Still the numerical estimation is a challenge (work in progress)
although many investigations are performed but not yet extensively for the FWI.

L2L0 regression $\ell^2\ \ell^0$ or L1L1 regression $\ell^1\ \ell^1$

$$\mathcal{C} = \mathcal{C}_{data} + \mathcal{C}_{model} = \frac{1}{2}||\Delta d||^2_{W_d} + \frac{1}{2}\epsilon||\delta m||^0_{W_m}$$

$$\mathcal{C} = \mathcal{C}_{data} + \mathcal{C}_{model} = \frac{1}{2}|\sqrt{W_d}\Delta d| + \frac{1}{2}\epsilon|\sqrt{W_m}\delta m|$$

Not yet investigated as far as we know because problem of convexity ...
Should be strongly connected with the problem to be solved ...

Model perturbation ℓ^2 ℓ^2

Two identical formulations depending on the availability of W_m^{-1} or W_m

$$\mathcal{C}_{i+1} = \frac{1}{2}((d^{obs} - u(m_i))^t W_d (d^{obs} - u(m_i))^*) + \frac{1}{2}\epsilon((m_i - m_{prior})^t W_m (m_i - m_{prior})^*)$$
$$\delta m_{i+1} = \mathcal{R}\left[J_i^t W_d J_i^* + \epsilon W_m\right]^{-1} \mathcal{R}\left[J_i W_d (d^{obs} - u(m_i))^* + \epsilon W_m (m_i - m_{prior})\right]$$

$$\mathcal{C}_{i+1} = \frac{1}{2}((d^{obs} - u(m_i))^t W_d (d^{obs} - u(m_i))^*) + \frac{1}{2}\epsilon((m_i - m_{prior})^t W_m (m_i - m_{prior})^*)$$
$$\delta m_{i+1} = \mathcal{R}\left[W_m^{-1} J_i^t W_d J_i^* + \epsilon I\right]^{-1} \mathcal{R}\left[W_m^{-1} J_i W_d (d^{obs} - u(m_i))^* + \epsilon (m_i - m_{prior})\right]$$

In current FWI, the second formulation is used
with a damping and smoothing in both gradient term and Hessian term using $m_{prior} = m_i$

$$\mathcal{C}_{i+1} = \frac{1}{2}((d^{obs} - u(m_i))^t W_d (d^{obs} - u(m_i))^*)$$
$$\delta m_{i+1} = \mathcal{R}\left[W_m^{-1} J_i^t W_d J_i^* + \epsilon I\right]^{-1} \mathcal{R}\left[W_m^{-1} J_i W_d (d^{obs} - u(m_i))^*\right]$$

Work in progress in this direction of better tuning the a priori information and the smoothing of your medium

Outline

Conclusion

- Currently, simple formulation of the inversion as the forward problem is quite intensive
- On the side of data space
  - Different levels of hierarchy in the data processing
  - Frequency hierarchy from low to high frequencies
  - Time windowing from EWT to FWT
  - Slant-stack hierarchy
- On the side of model space
  - Initial model definition
  - A priori model influence
  - A priori uncertainties and correlation lengths estimation
- On the side of objective function
  - Various formulations of objective functions
  - cross-correlation (Tape et al., 2009; Baumstein et al., 2011)
  - time-frequency criteria (Fichtner et al., 2009)
  - Improved Hessian formulation
  - Hyper-parameters tuning : weighting operators

Conclusion

FWI is now going into a mature level with many applications for a better tuning of various parameters.

FWI may provide a macromodel inside which migration will focus on well identified phases.

Thanks you for your attention

Backus, G. and Gilbert, J. F. (1967). Numerical applications of a formalism for geophysical inverse problems. *Geophys. J. R. Ast. Soc.*, 13 :247–276.

Baumstein, A., Ross, W., and Lee, S. (2011). Simultaneous source elastic inversion of surface waves. In *Expanded Abstracts*, page C040. European Association of Geoscientists and Engineers.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2009). Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3) :1703–1725.

Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveform : numerical results. *Geophysics*, 51 :1387–1403.

Mora, P. R. (1987). Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52 :1211–1228.

Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151) :773–782.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. New York, US : Springer.

Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2) :495–503.

Polak, E. and Ribière, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *Revue Française d'Informatique et de Recherche Opérationnelle*, 16 :35–43.

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

Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2009). Seismic tomography of the southern california crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180 :433–462.

Tarantola, A. (1984). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32 :998–1015.

Tikhonov, A. and Arsenin, V. (1977). *Solution of ill-posed problems*. Winston, Washington, DC.