Full Waveform Inversion in Practice

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

12th International Congress of the Brazilian Geophysical Society
Short Course on Full Waveform Inversion
15 August 2011
1 Introduction
2 FWI Resolution
3 Multiscale algorithms
4 Source time function estimation
5 Starting model building
This part of the course is mainly based on the following paper and presentation


1. Introduction
2. FWI Resolution
3. Multiscale algorithms
4. Source time function estimation
5. Starting model building
FWI principle
Intuitive understanding in time domain ... the incident wavefield
FWI principle

Intuitive understanding in time domain ... the adjoint wavefield from receivers

Time: 3.5 s
FWI principle
Correlation of the incident and adjoint wavefields
Key questions to be addressed

1. What is the FWI resolution?
   - Reminders on the diffraction tomography
   - Effects of the Hessian information

2. Multiscale/Hierarchical FWI algorithms

3. Source time function reconstruction?

4. Starting model definition?
   - Accuracy and cycle skipping
   - Possible methods for building starting models
Outline

1. Introduction

2. FWI Resolution
   - FWI gradient estimation
   - FWI Hessian estimation

3. Multiscale algorithms

4. Source time function estimation

5. Starting model building
Outline

1 Introduction

2 FWI Resolution
   - FWI gradient estimation
   - FWI Hessian estimation

3 Multiscale algorithms

4 Source time function estimation

5 Starting model building
Monochromatic gradient: The Fresnel zone

\[ \theta \]

\[ S \]

\[ R \]

\[ k \]

Distance (km)
Depth (km)

\[ 0 \quad 10 \quad 20 \quad 30 \quad 40 \quad 50 \quad 60 \quad 70 \quad 80 \quad 90 \quad 100 \]

\[ 0 \quad 5 \quad 10 \quad 15 \quad 20 \]

Distance (km)

\[ 0 \quad 1 \quad 2 \quad 3 \quad 4 \]

Distance (km)

\[ 0 \quad 1 \quad 2 \quad 3 \quad 4 \]

Distance (km)

\[ 0 \quad 1 \quad 2 \quad 3 \quad 4 \]

Distance (km)

\[ 0 \quad 1 \quad 2 \quad 3 \quad 4 \]
Relation between frequency, aperture and resolution

- FWI is based on diffraction tomography principle and can be seen as a spatial inverse Fourier transform (Devaney, 1982)
- Spatial resolution of FWI (cf. Sirgue and Pratt, 2004)

\[ \vec{k} = \frac{2\omega}{c_0} \cos \left( \frac{\theta}{2} \right) \vec{n}, \]  

(1)

- \( \vec{k} \): wavenumber for the acquisition aperture
- \( f \): frequency for the acquisition time series
Relation between frequency, aperture and resolution

- FWI is based on diffraction tomography principle and can be seen as a spatial inverse Fourier transform (Devaney, 1982)
- Spatial resolution of FWI (cf. Sirgue and Pratt, 2004)

\[ \vec{k} = \frac{2\omega}{c_0} \cos(\theta/2) \hat{n}, \]  

(2)
Impact on acquisition
Transmission acquisition: single ball

7 freqs, 5 iter/freq, [4, 5, 7, 10, 13, 16, 20] Hz
Impact on acquisition

Transmission acquisition: single ball (2)

7 freqs, 5 iter/freq, [4, 5, 7, 10, 13, 16, 20] Hz
Impact on acquisition
Transmission + reflection acquisition: single ball

7 freqs, 5 iter/freq, [4, 5, 7, 10, 13, 16, 20] Hz
Impact on acquisition

Transmission acquisition: two balls

7 freqs, 5 iter/freq, [4, 5, 7, 10, 13, 16, 20]Hz
Impact on acquisition
Ray based tomography
Impact on acquisition
Transmission + reflection acquisition: two balls

7 freqs, 5 iter/freq, [4, 5, 7, 10, 13, 16, 20] Hz
Impact on acquisition

a last example

- Source in red
- Receiver in blue
Resolution and wavenumbers

Conclusion

To avoid local minima, the full spatial wavenumber must be filled
FWI requires wide angle acquisitions
  - Large angle (long offsets) to constraint low wavenumber (long wavelengths)
  - Small angles (short offset) to reach the optimal resolution ($\lambda/2$)
Thanks to redundancy, the frequency spectrum could be decimated
Outline

1. Introduction

2. FWI Resolution
   - FWI gradient estimation
   - FWI Hessian estimation

3. Multiscale algorithms

4. Source time function estimation

5. Starting model building
The Hessian (related to resolution matrix) contains information on spatial correlation between diffractors.

Using Hessian in optimization introduces this information: acts as a deconvolving operator on the gradient.
The Hessian (related to resolution matrix) contains information on spatial correlation between diffractors.

Using Hessian in optimization introduces this information: acts as a deconvolving operator on the gradient.

→ Gradient versus Gauss Newton optimization 4 freqs, 1 iter/freq, [4, 5, 7, 10]Hz.
The Hessian information improve significantly the results...

... but computationally expensive. Only few realistic applications with Newton or Gauss-Newton methods

Alternatives

▶ Quasi-Newton approach as L-BFGS (Brossier et al., 2009)
▶ Incomplete Gauss-Newton (Hu et al., 2009)
Outline

1 Introduction
2 FWI Resolution
3 Multiscale algorithms
4 Source time function estimation
5 Starting model building
Efficient frequency-domain FWI algorithms use only few discrete frequencies from low to high frequencies for wide aperture acquisitions (Sirgue and Pratt, 2004; Brenders and Pratt, 2007a)

- multiscale approach and mitigation of non-linearities
- efficiency

### Algorithm

1. **for** frequency = frequency\(_1\) to frequency\(_n\) **do**
2. \hspace{0.5cm} **while** (NOT convergence AND iter < niter\(_{\text{max}}\)) **do**
3. \hspace{1.5cm} Estimate source wavelet if required
4. \hspace{1.5cm} Build gradient vector \( \mathcal{G}_m^{(k)} \)
5. \hspace{1.5cm} Build perturbation vector \( \delta m \)
6. \hspace{1.5cm} Update model \( m^{(k+1)} = m^{(k)} + \alpha \delta m \)
7. \hspace{1.5cm} **end while**
8. **end for**
Efficient multiscale frequency-domain FWI algorithm

Synthetic application from Sourbier et al. (2009) (true, starting, 3.5 Hz, 9.2 Hz, 20.6 Hz)
Crustal imaging on the Nankai thrust, Japan (Operto et al., 2006)
Efficient multiscale frequency-domain FWI algorithm

Complex frequency preconditioning allows to select arrivals from the first arrival (Shin et al., 2002; Brenders and Pratt, 2007b).

- Remove complex late arrivals
- An heuristics to select apertures in the data

\[
F(\omega + i\gamma)e^{\gamma t_0} = \int_{-\infty}^{+\infty} f(t)e^{-\gamma(t-t_0)}e^{-i\omega t} \, dt
\]  

(3)
Two-levels hierarchical algorithm (Brossier et al., 2009)

1: \textbf{for} frequency $=$ $\text{frequency}_1$ to $\text{frequency}_n$ \textbf{do}
2: \hspace{1em} \textbf{for} data
damping $=$ $\text{high}_d$amping to $\text{low}_d$amping \textbf{do}
3: \hspace{2em} \textbf{while} (NOT convergence AND iter $<$ $\text{niter}_{\text{max}}$) \textbf{do}
4: \hspace{3em} Estimate source wavelet if required
5: \hspace{3em} Build gradient vector $G_m^{(k)}$
6: \hspace{3em} Build perturbation vector $\delta m$
7: \hspace{3em} Update model $m^{(k+1)} = m^{(k)} + \alpha \delta m$
8: \hspace{2em} \textbf{end while}
9: \hspace{1em} \textbf{end for}
10: \textbf{end for}
Outline

1. Introduction
2. FWI Resolution
3. Multiscale algorithms
4. Source time function estimation
5. Starting model building
The source time function can be estimated as an unknown of the inverse problem (Pratt, 1999):

\[ C = \frac{1}{2} \Delta d^\dagger \Delta d \]  

- \( \Delta d = d_{\text{obs}} - d_{\text{calc}} \)
- \( d_{\text{calc}} = sg \) with \( g \) the Green function

\[ C = \frac{1}{2} (d_{\text{obs}} - sg)^\dagger (d_{\text{obs}} - sg) \]  

\[ s = \frac{\langle g_{\text{cal}}(m_0) | d_{\text{obs}} \rangle}{\langle g_{\text{cal}}(m_0) | g_{\text{cal}}(m_0) \rangle} , \]  

It can be seen as the division of the cross-correlation of the computed and observed data, by the autocorrelation of the computed data.
Source time function estimation

- Analysis of the repeatability of source wavelet to qualify model accuracy (under the assumption that sources are repeatable on the field)

![Graphs showing different models: NMO, FATT, Anisotropic, FWI, Anisotropic FWI, 3D FWI, NMO+FWI, FATT+FWI model 1]
Current multidimensional FWI based on local optimization
- High number of unknowns
- Numerical cost of forward problem

For classical misfit functions ($L_2$, $L_1$, log,...), local method implies a “good” starting model
- No cycle skipping of the data
- Wavenumber content close to the smallest wavenumbers of FWI (to avoid gap in the spectrum)
Limitations for starting from scratch
  ▶ Limited low frequencies in the data
    (recent studies ≈ 1-2 Hz, Plessix et al., 2010)
  ▶ Incomplete illumination
Limitations for starting from scratch
  ▶ Limited low frequencies in the data
    (recent studies $\approx$ 1-2 Hz, Plessix et al., 2010)
  ▶ Incomplete illumination

An other method is required to delineate the starting model of FWI
Possibles methods and their limitations
Reflection tomography and MVA

- Pros
  - Usage of reflected events
  - Strong experience in O&G industry
  - High resolution for FWI

- Cons
  - Applicability in complex environments?
  - Sensitive to NMO velocity: validity at long offset in case of anisotropy?
    (discussion later on the anisotropy issue)

- Examples in Sirgue et al. (2009, 2010)

Images from Sirgue et al. (2009)
Possibles methods and their limitations
First-arrival travel times tomography

Pros

▶ Robust
▶ Weak numerical cost (in particular with adjoint formulation Taillandier et al., 2009)

Cons

▶ Requires picking of first break
▶ Low resolution (requires low frequency for FWI)
▶ Limited penetration in case of LVZ
▶ Sensitive to “horizontal” velocity

Some examples in Ravaut et al. (2004); Operto et al. (2006); Brenders and Pratt (2007b,a)
Possibles methods and their limitations

Stereotomography

Pros

▶ Can use reflected and refracted events (slope of locally coherent events)
▶ Hierarchical scheme from large to weak aperture (Prieux et al., 2010)

Cons

▶ Requires picking (automatic) and QC
▶ Numerical cost of the hierarchical scheme

preliminary example of Stereotomo + FWI in Prieux et al. (2010), review of stereotomo in Lambaré (2008)
Possibles methods and their limitations

Laplace domain inversion

- **Pros**
  - Seems efficient on large contrast (salt)
  - Affordable cost

- **Cons**
  - Requires picking of the first break and careful mute to ensure robustness
  - Requires long time recordings
  - Instabilities (incomplete Laplace transforms)

- Example in Shin and Ha (2009)

(image courtesy of C. Shin and Y. H. Cha)
Possibles methods and their limitations

Early waveform tomography (EWT)

- Avoid phase ambiguity by considering only first phases
- Increase the size of the minimum valley
- Expected better resolution than traveltime tomography Sheng et al. (2006); Ellefsen (2009)

left (FWI), middle (EWT), right (travel time) (Sheng et al., 2006)
Currently, no generic method for starting model building...

The choice in mainly driven by the data properties (acquisition, frequency-contain...)

... still an active research field

Some alternative choice of the misfit function are proposed to avoid the the cycle-skipping limitation

- cross-correlation (Tape et al., 2009; Baumstein et al., 2011)
- time-frequency criteria (Fichtner et al., 2009)


