PART III

ITERATIVE SOLUTION OF THE TOMOGRAPHIC INVERSE PROBLEM

1. French cheese, non-linearity and the multi-scale approach
2. Misfit functionals
3. Towards quantitative resolution analysis
French cheese, non-linearity and the multi-scale approach
The Camembert experiment (Gauthier, Virieux & Tarantola, 1986)

- Synthetic full waveform inversion experiment.
- The misfit was the most obvious $L_2$ difference:

$$X(m) = \frac{1}{2} \int [u - u_0]^2 \, dt$$

- The goal was to recover an input model with the shape of a Camembert.
The Camembert experiment (Gauthier, Virieux & Tarantola, 1986)

- Synthetic full waveform inversion experiment.
- The misfit was the most obvious $L_2$ difference:

$$X(m) = \frac{1}{2} \int [u - u_0]^2 \, dt$$

- The goal was to recover an input model with the shape of a Camembert.
The Camembert experiment (Gauthier, Virieux & Tarantola, 1986)

The inversion was trapped in a local minimum of the misfit function.

- Seismic waveforms depend non-linearly on the structure of the Earth.
- This can result in the presence of multiple local minima of the misfit.
Non-linearity and multiple minima

1. Start from initial Earth model $m_0$

2. Update according to $m_{i+1} = m_i + \gamma_i h_i$, with $\chi(m_{i+1}) < \chi(m_i)$

Diagram:
- $\chi$ axis
- $m_0$ and $m_1$ points
- $\gamma_0 h_0$ step length
- Descent direction
1. Start from initial Earth model \( m_0 \)

2. Update according to \( m_{i+1} = m_i + \gamma_i h_i, \) with \( \chi(m_{i+1}) < \chi(m_i) \)

Iteratively approach the minimum misfit by following the local descent directions.
Non-linearity and multiple minima

1. Start from initial Earth model \( m_0 \)

2. Update according to \( m_{i+1} = m_i + \gamma_i h_i \), with \( \chi(m_{i+1}) < \chi(m_i) \)

![Graph showing the iterative update process and non-linearity with multiple minima.](image)

- step length
- descent direction

\( m_0 \) \( m_1 \) \( m_2 \) \( m_3 \) \( m_4 \) \( \ldots \)
Non-linearity and multiple minima

1. Start from initial Earth model $\mathbf{m}_0$

2. Update according to $\mathbf{m}_{i+1} = \mathbf{m}_i + \gamma_i \mathbf{h}_i$, with $\chi(\mathbf{m}_{i+1}) < \chi(\mathbf{m}_i)$

![Diagram showing iterative updates with non-linearity and multiple minima]
Gradient methods are **local**.

Convergence to the **global minimum** relies on a **good initial model**.

Good initial model: e.g. long-wavelength model from ray tomography.

Bleibinhaus et al., 2007
Graduate methods are local.

Convergence to the global minimum relies on a good initial model.

Good initial model: e.g. long-wavelength model from ray tomography.

Bleibinhaus et al., 2007
Sufficiently good initial models are often not available.

The multi-scale approach is an empirical strategy that helps to overcome this problem.
The multi-scale approach

long-period data → long-wavelength structure
The multi-scale approach

shorter-period data $\rightarrow$ shorter-wavelength structure
The multi-scale approach

\[ \chi(m) \]

1. stage

\[ m_0 \quad \tilde{m}_1 \]

2. stage

\[ \tilde{m}_1 \quad \tilde{m}_2 \]

3. stage

\[ \tilde{m} \]

Increasing detail

short-period data

→ short-wavelength structure
The multi-scale approach

\[ \chi(m) \]

1. stage

\[ m_0 \]

\[ \tilde{m}_1 \]

2. stage

\[ \chi(m) \]

\[ m_2 \]

3. stage

\[ \chi(m) \]

\[ m \]

increasing detail

Salinian distance [km]

Franciscan

Great Valley Seq.

SW

10

15

20

25

30

35

40

45

traveltime

4 Hz

8 Hz

12 Hz

p-wave velocity [km/s]

2.4

3.2

4.0

4.8

5.6

6.4

Bleibinhaus et al., 2007
Misfit functionals
A misfit functional suitable for waveform tomography should:

1) extract as much information as possible from seismic waveforms

2) be quasi-linearly related to Earth structure (tame non-linearity)
A misfit functional suitable for waveform tomography should:

1) extract as much information as possible from seismic waveforms

2) be quasi-linearly related to Earth structure (tame non-linearity)

"P.S.: Do you use amplitude information? If so, be really careful with the temporary data. I am prepared to bet that you can have amplitude errors of a factor of up to 10 …

This type of caution even applies for many permanent observatory stations. I could show you an interesting figure on that subject, which shows that approximately 40% of permanent broad-band stations have amplitude errors >10% …"
A misfit functional suitable for waveform tomography should:

1) extract as much information as possible from seismic waveforms
2) be quasi-linearly related to Earth structure (tame non-linearity)
3) be independent of absolute amplitudes
Synthetic experiment to study the detectability of mantle plumes.

Plume model (d=500km, -5% P wave speed, -10% S wave speed).

Idealistic source-receiver geometry.

Measurement of cross-correlation time shifts in 30 s P waves.
Example 1: Hunting for plumes in the mantle (Florian Rickers)

- Misfit reduction after 7 conjugate-gradient iterations: 95% 
- Traveltimes are essentially explained.
- The plume remains diffuse and is not visible below 600 km.
Example 1: Hunting for plumes in the mantle (Florian Rickers)
Design of misfit functionals

Example 1: Hunting for plumes in the mantle (Florian Rickers)

- Traveltime differences decay rapidly due to wavefront healing.
- Deeper parts of the plume are not detectable by cross-correlation.
Design of misfit functionals

Example 1: Hunting for plumes in the mantle (Florian Rickers)

- Plume acts as a scatterer.
- Away from the direct ray path, plume information arrives in the coda of the P-wave.
- This information is not accessible by cross-correlation.
Design of misfit functionals

Example 1: Hunting for plumes in the mantle (Florian Rickers)

Alternative measurement: instantaneous phase differences of the complete seismogram

- analytic signal: \( \tilde{f}(t) = f(t) + i Hf(t) \)
- instantaneous phase: \( \Phi(t) = \arctan \frac{\text{Im} \tilde{f}}{\text{Re} \tilde{f}} \)
- instantaneous phase misfit: \( \chi = \sum_{i=1}^{N_r} \int [\Phi_i^0(t) - \Phi_i(t)]^2 dt \)

Advantages:

- Considers the complete time series, including the scattered waves.
- All the „wiggles“ are automatically weighted equally.
- Independent of absolute amplitudes.
Example 1: Hunting for plumes in the mantle (Florian Rickers)

- Instantaneous phase inversion for the synthetic plume.
- Use the P wave train until the arrival of the S wave.
- Misfit reduction of 90% after 9 iterations.
- Much sharper image of the plume (... but still not to great depth).

Cross correlation traveltime

Instantaneous phase

Design of misfit functionals
Design of misfit functionals

Example 2: Time-frequency misfits

Time-frequency misfits
phase differences as functions of time and frequency
Time-frequency misfits

phase differences as functions of time and frequency

- quasi-linearly related to Earth structure
  improves convergence

- independent of amplitudes
  reliably measurable, deep structure information

- applicable to complex waveforms
  interfering waves, unidentifyable waves

- continuous in frequency
  no discrete frequency bands
Design of misfit functionals

Example 2: Time-frequency misfits

isotropic S wave speed
Towards quantitative resolution analysis
Resolution analysis

How can we quantify the trustworthiness of FWI images?

Arguments typically involve:

1. visual analysis
2. synthetic inversions

Are FWI images really any better?

Is FWI worth the effort?
Resolution analysis

Complications: non-linearity, multiple minima, finite computational resources, …

Quadratic approximation of the misfit functional

\[ \chi(m_{opt} + \delta m) \approx \chi(m_{opt}) + \delta m^T H \delta m \]

The Hessian \( H \):

1. local geometry of the misfit surface
2. resolution and trade-offs
3. \( H = \text{inverse posterior covariance} \)
Efficient computation of $H \cdot \delta m$ via an extension of the adjoint method

1. 2 forward simulations
2. 2 adjoint simulations (time-reversed)
Efficient computation of $H \cdot \delta m$ via an extension of the adjoint method

1. 2 forward simulations
2. 2 adjoint simulations (time-reversed)

Example: 25 s Love wave
finite-frequency traveltime

Fréchet kernel

Efficient computation of $H \cdot \delta m$ via an extension of the adjoint method

1. 2 forward simulations
2. 2 adjoint simulations (time-reversed)

Example: 25 s Love wave
finite-frequency traveltime

$\delta m = v_s$ perturbation in pixel $k$

Efficient computation of $H \cdot \delta m$ via an extension of the adjoint method

1. 2 forward simulations
2. 2 adjoint simulations (time-reversed)

Example: 25 s Love wave
finite-frequency traveltime

$\delta m = v_s$ perturbation in pixel $k$

Resolution analysis

$F$: approximate Hessian (first-order scattering)
$S$: second-order scattering

Resolution analysis

Long-wavelength FWI image of Europe: **EU-MOD.ISO.3**

- **Data:** ≈ 3000 waveforms
  - dominant period: 100 s
- **3 iterations**
Example: A Hessian column (kernel) for northern Germany

Ideally, we would like to have such a Hessian kernel for each point in the model.

This would be a computationally expensive exercise, but ...
Resolution analysis

\[ H \cdot \exp(i k \cdot x) = \text{Fourier transform of the H-columns} \]

- approximate \( H \) from a small number of wave number vectors \( k \)
- estimate width of the **influence length** \( \ell \) in different directions

The diagram illustrates the effect of the influence length \( \ell_x \) on pixel \( j \). There is **no effect** on pixel \( j \) outside the influence length, and an **effect** on pixel \( j \) within the influence length.
Resolution analysis

\[ H \cdot \exp(i k \cdot x) = \text{Fourier transform of the H-columns} \]

- approximate \( H \) from a small number of wave number vectors \( k \)
- estimate width of the influence length \( l \) in different directions
Resolution analysis

- $H \cdot \exp(ik \cdot x) = \text{Fourier transform of the } H\text{-columns}$
- approximate $H$ from a small number of wave number vectors $k$
- estimate width of the influence length $\ell$ in different directions

good estimate of $H$ from a hand full of forward and adjoint simulations
Resolution analysis

1. extension of the adjoint method to compute $H \cdot \delta m$

2. $H \cdot \delta m =$ superposition of approximate Hessian and second-order scattering

3. resolution and influence region of geologic structures (e.g. Iceland plume)

4. trade-offs between parameters (e.g. $v_s$, $v_p$, $\rho$)

5. spectral estimation of the Hessian through Fourier transforms

6. continuous distribution of the influence length

First step towards quantitative resolution analysis in full waveform inversion.
Lorentz Center Workshop on

Uncertainty Analysis in Geophysical Inverse Problems

7 – 11 November 2011
Leiden, The Netherlands