Le Camembert Fondu: Albert Tarantola and the dawn of Full Waveform Inversion

William Symes

The Rice Inversion Project

SEG, 2010

Agenda

Inverse Problem Theory

A milestone paper

That was then, this is now...

Summary

A Pair of Major Contributions

- Probabilistic inverse theory - Bayesian inference of physical quantities: *Inverse Problem Theory* (1987) and many papers before and after

  > *"Realistic inverse problems are, generally, nonlinear. When nonlinearities are strong, deterministic, iterative methods are not useful....", web intro to* Monte Carlo Sampling of Solutions to Inverse Problems, *Mosegaard and Tarantola, JGI 1995.*

- Deterministic, iterative methods: practical descent methods, milestone numerical explorations

Agenda

Inverse Problem Theory

A milestone paper

That was then, this is now...

Summary

Gauthier et al., Geophysics *51* (1986)

GEOPHYSICS, VOL. 51, NO. 7 (JULY 1986); P. 1387–1403, 16 FIGS.

Two-dimensional nonlinear inversion of seismic waveforms: Numerical results

Odile Gauthier*, Jean Virieux*, and Albert Tarantola*

ABSTRACT

The nonlinear problem of inversion of seismic waveforms can be set up using least-squares methods. The inverse problem is then reduced to the problem of minimizing a (nonquadratic) function in a space of many (10^4 to 10^6) variables. Using gradient methods leads to iterative algorithms, each iteration implying a forward propagation generated by the actual sources, a backward propagation generated by the data residuals (acting as if they were sources), and a correlation at each point of the space of the two fields thus obtained, which gives the updated model. The quality of the re-

If a starting model is known to be close enough to the actual medium, linearized inversion will probably perform well. Unfortunately, since there is no practical test to check the accuracy of the linearization, nonlinear inverse techniques are more promising. There are two basic options: a full exploration of the parameter space (systematic or Monte Carlo), or a local descent method. The first approach has the advantage of avoiding local minima, but it is too time-consuming for modern computers. The second approach gives the correct solution when the starting model is inside the valley of the global minimum, irrespective of the choice of starting point. Thus the starting model plays a much less crucial role than in linearized inversion.

Everything you wanted to know about FWI, but were afraid to ask... and it's reproducible!

Gauthier et al., Geophysics *51* (1986)

This paper ("GVT"): first published exploration of iterative FWI with multi-D data *and* multi-D models

It's all here:

- adjoint state method for gradient computation
- reflection is hard
- transmission is easier, but nontrivial
- key issue: interaction of spectral data incompleteness with strong nonlinearity

Adjoint State Gradient Computation

F = *forward map* or *modeling operator*, model $m \mapsto$ data $F[m]$

For Bayesian or other reasons, want to compute

$$m_{\mathrm{ML}} = \mathrm{argmin}\left[(F[m]-d)^T C_d (F[m]-d)\right.$$

$$\left.+(m-m_{\mathrm{prior}})^T C_m (m-m_{\mathrm{prior}})\right]$$

Large scale $\Rightarrow$ need fast convergence $\Rightarrow$ Newton $\Rightarrow$ need $DF[m]^T$

Practical issue: F computed by recursive / iterative process - how to compute $DF[m]^T$ with similar economy?

Adjoint State Gradient Computation

Adjoint state method: recursive computation of $DF[m]^T$ - computational complexity comparable to that of $F[m]$

- bottom line: $(A_1 A_2 ... A_n)^T = A_n^T ... A_2^T A_1^T$
- optimal control theory - adjoint field $\sim$ costate, Lagrange multiplier - Pontryagin Maximum Principle
- optimal control of PDEs - J.-L. Lions et al., 60's
- introduced into computational inverse problems - Chavent & Lemmonier, 1974
- inverse problems for (1D) waves - Bamberger, Chavent & Lailly 1977, 1979

Adjoint State Gradient Computation

- relation to prestack RTM - Tarantola, Lailly early 80's
- Tarantola *Geophysics* v. 49 (1984) - explanation of the method for acoustics

- Tarantola IPG group - many applications to multi-D inversion, synthetic and field data, 80's & early 90's
  - Cao et. al 1990: acoustic, marine 2D
  - Crase et al 1990: elastic, marine 2D
  - Mora's thesis
  - Jananne et al. 1989: resolvable wavelengths
  - many others ...

GVT = first full-blown application with multi-D data, model.

The Camembert

Principal example in GVT

- 1 km $\times$ 1 km domain
- background: v $= 2.5$ km/s, $\rho = 4$ g/cm^3
- peak frequency ~ 50 Hz
- circular bulk modulus anomaly, diam $= 1$ km, "small" $= 2\%$ or "large" $= 20\%$
- "nonlinear saturation" at 10% - full wavelength traveltime perturbation

The Camembert

- 5 m grid - 10 gridpts / (peak) wavelength
- absorbing BCs on all sides
- reflection configuration:
  - 100 receivers (fixed spread), $z_r = 80, x_r = 0, 10, ..., 990$ m
  - 8 sources, $z_s = 40, x_s = 110, 220, ..., 880$ m
- transmission configuration:
  - 400 receivers - each side like top in reflection
  - 8 sources at corners and side midpoints

The Camembert

1392 Gauthier et al.

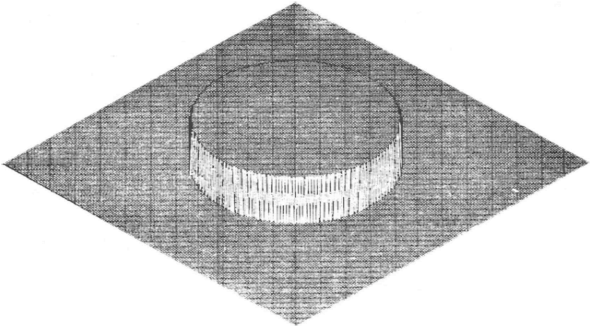


FIG. 5. The model is now a circular inclusion (the "Camembert") in a homogeneous medium. The size of the Camembert is about 10 wavelengths. The model is numerically defined in a grid with 200 × 200 points, so the model contains 10^4 parameters (unknowns for the inversion).

The Camembert

Reproduction of Camembert tests using IWAVE++: FWI, based on IWAVE modeling engine (thanks: Dong Sun)

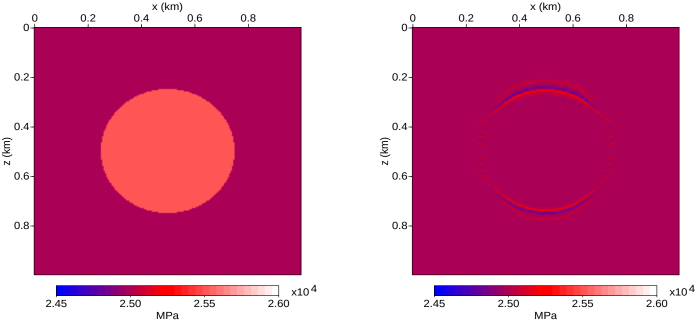
Staggered grid acoustic FD scheme after Virieux 1984, many others.

A few differences:

-	GVT	IWAVE++
ABC	E-M	PML
Scheme	2-2	2-4
Src fld	save bdry	opt chkpt
Opt Alg	NLCG	LBFGS

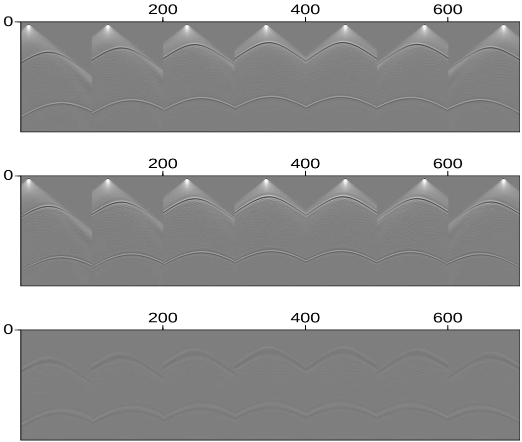
Small Anomaly - reflection

Initial MS resid = 3629; Final after 5 LBFGS steps = 254



Bulk modulus: Left, model; Right, inverted

Small Anomaly - reflection



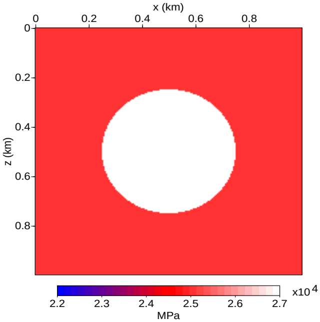
Top: target data; Middle: estimated data; Bottom: difference

Small Anomaly - reflection

Message: "the Camembert has melted". Linear ambiguity (ill-conditioning) from spectral incompleteness

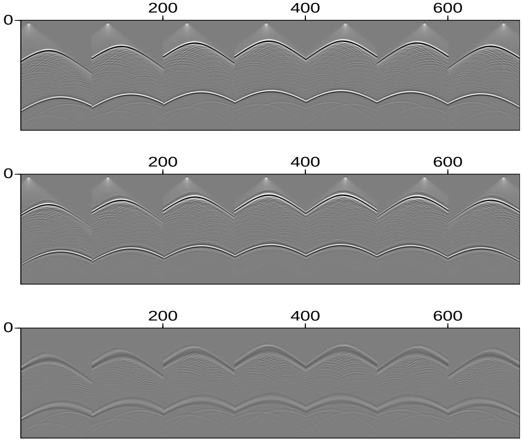
Large Anomaly - reflection

Initial MS resid $= 301 \times 10^3$; Final after 5 LBFGS steps $= 24 \times 10^3$



Bulk modulus: Left, model; Right, inverted

Large Anomaly - reflection



Top: target data; Middle: estimated data; Bottom: difference

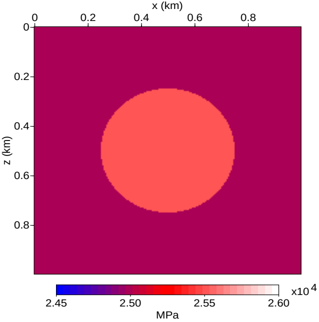
Large Anomaly - reflection

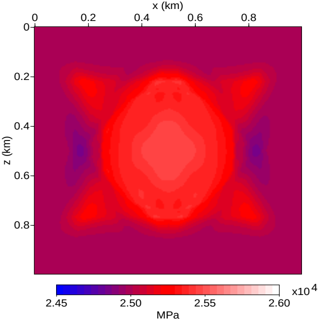
Message: no evidence of cycle skipping, traveltime mismatch of primary reflections: failure to recover velocity macromodel *still mostly* linear ambiguity from spectral incompleteness, even though problem is “saturated”.

However prismatic multiple reflections not fit - possibly near secondary min.

Small Anomaly - transmission

Initial MS resid $= 2.56\times10^7$; Final after 5 LBFGS steps $= 2.6\times10^5$





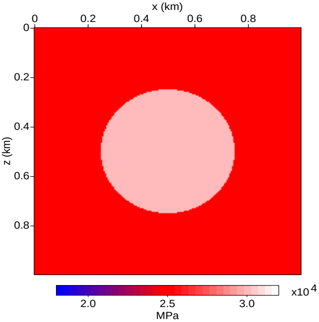
Bulk modulus: Left, model; Right, inverted

Small Anomaly - transmission

Message: transmission is much better conditioned than reflection in linear regime, overcomes spectral incompleteness of data

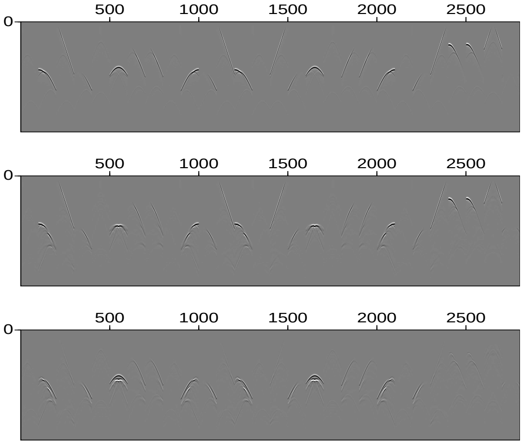
Large Anomaly - transmission

Initial MS resid = 2.14×10^8; Final after 5 LBFGS steps = 1.44×10^8



Bulk modulus: Left, model; Right, inverted

Large Anomaly - transmission



Top: target data; Middle: estimated data; Bottom: difference

Large Anomaly - transmission

Message: transmission is very sensitive to nonlinear saturation - cycle-skips evident - likely close to large-residual local min

Agenda

Inverse Problem Theory

A milestone paper

That was then, this is now...

Summary

Reflection

Still the “hard case” - subsequent work $\Rightarrow$ low residual local mins (spectral incompleteness) may coexist with high residual local mins

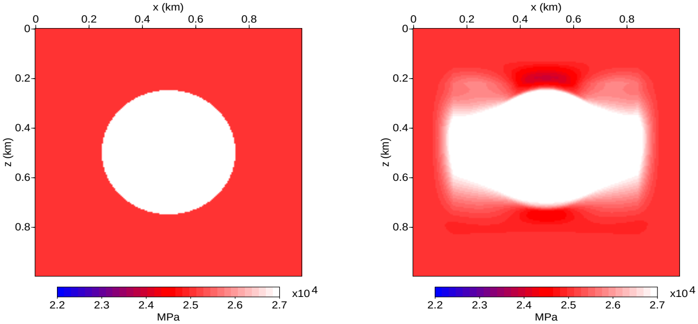
Conventional remedies:

- start close - use traveltime tomography or velocity analysis to devise initial model. Q: how close?
- continuation in depth, time, frequency - pioneered by Lailly group (Kolb et al. 1986), many others. Q: adequate starting depth/time/frequency?
- more complete spectrum, low data frequencies - Bunks et al. 1995, many others. Q: where do you get them?

Reflection

Large anomaly - reflection - impulsive data, 0-60 Hz

Initial MS resid $= 3.45\times10^{13}$; Final after 5 LBFGS steps $=$ 4.78×10^{11}



Bulk modulus: Left, model; Right, inverted

Transmission

Pioneering work by Pratt group (Pratt 1999, Pratt & Shipp 1999, Sirgue & Pratt 2004, Brenders & Pratt 2006) exploits "transmission is easier (but not *easy*)":

- continuation, decimation in frequency
- continuation in time (windows, exponential weights)
- formulation in frequency domain

Many recent examples of strikingly effective surface-data FWI using diving wave energy - see FWI sessions

And So On...

GVT p. 1395:

> *The solution of the problem will probably be found through better use of the traveltime information in the data set.*

Two approaches:

- traveltime tomography (transmission, reflection, stereo-) as preprocess, to establish initial model - many examples in this week's FWI sessions
- MVA $\sim$ solution approach to Born FWI - can MVA concepts (image gathers, semblance measures) be imported into FWI?

Agenda

Inverse Problem Theory

A milestone paper

That was then, this is now...

Summary

RICE

Summary

GVT paper:

- presents first exposition of FWI with multi-D data & models
- anticipates much of current state of knowledge
- illustrates AT's insight and scientific integrity

Summary

thanks: J. Claerbout

Summary

Thanks to

LH and WB for invitation to speak

Dong Sun for recreating the Camembert examples with IWAVE++

Present and former TRIP team members: Dong Sun, Igor Terentyev, Tanya Vdovina, Rami Nammour, Marco Enriquez, Xin Wang, Chao Wang

Sponsors of The Rice Inversion Project

Reconstructed Camembert data - IWAVE demo package, http://www.trip.caam.rice.edu/software