Objective functions for full waveform inversion

William Symes

The Rice Inversion Project

EAGE 2012

Workshop: From Kinematic to Waveform Inversion

A Tribute to Patrick Lailly

Agenda

Overview

Challenges for FWI

Extended modeling

Summary

Full Waveform Inversion

M = model space, D = data space

$F : M \to D$ forward model

Least squares inversion ("FWI"): given $d \in D$, find $m \in M$ to minimize

$$J_{LS}[m] = \|F[m] - d\|^2 [+ \text{ regularizing terms}]$$

($\|\cdot\|^2$ = mean square)

Full Waveform Inversion

- $+$ accommodates any modeling physics, data geometry, spatial variation on all scales (Bamberger, Chavent & Lailly 79,...)
- $+$ close relation to prestack migration via local optimization (Lailly 83, Tarantola 84)
- $+$ gains in hard/software, algorithm efficiency $\Rightarrow$ feasible data processing method
- $++$ some spectacular successes with 3D field data (keep listening!)
- $\pm$ with regularizations pioneered by Pratt and others, applicable surface data *if* sufficient (i) low frequency s/n and (ii) long offsets
- \- reflection data still a challenge

Full Waveform Inversion

Why are

- low frequencies important?
- long offsets (diving waves, transmission) easier than short offsets (reflections)?

What alternatives to Standard FWI = output least squares?

- different error measures, domains - time vs. Fourier vs. Laplace, L1, logarithmic - other talks today, survey Virieux & Operto 09
- model extensions - migration velocity analysis as a linearization, nonlinear MVA

Agenda

Overview

Challenges for FWI

Extended modeling

Summary

Nonlinear Challenges: Why low frequencies are important

Well-established observation, based on heuristic arguments ("cycle skipping"), numerical evidence :forward modeling operator is *more linear* [objective function is more quadratic] at lower frequencies

Leads to widely-used frequency continuation strategy (Kolb, Collino, & Lailly 86)

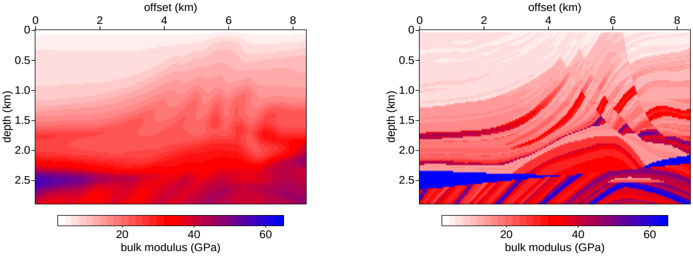
Why?

Nonlinear Challenges: Why low frequencies are important

Visualizing the shape of the objective: *scan* from model m_0 to model m_1

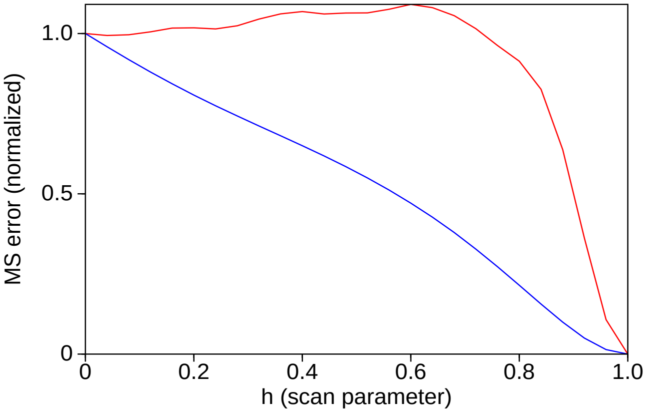
$$f(h) = J_{LS}[(1-h)m_0 + hm_1]$$

Expl: data = simulation of Marmousi data (Versteeg & Gray 91), with bandpass filter source.



m_0 = smoothed Marmousi, m_1 = Marmousi (bulk modulus displayed)

Nonlinear Challenges: Why low frequencies are important



Red: [2,5,40,50] Hz data. Blue: [2,4,8,12] Hz data

Nonlinear Challenges: Why low frequencies are important

Origin of this phenomenon in math of symmetric hyperbolic systems:

$$A\frac{\partial u}{\partial t} + Pu = f$$

$u =$ dynamical field vector, $A =$ symm. positive operator, $P =$ skew-symm. differential operator in space variables, $f =$ source

Example: for acoustics, $u = (p, \mathbf{v})^T$, $A = \mathrm{diag}(1/\kappa, \rho)$, and

$$P = \left(\begin{array}{cc} 0 & \mathrm{div} \\ \mathrm{grad} & 0 \end{array} \right)$$

Nonlinear Challenges: Why low frequencies are important

Theoretical development, including non-smooth A: Blazek, Stolk & S. 08, Stolk 00, after Bamberger, Chavent & Lailly 79, Lions 68.

Sketch of linearization analysis - after Lavrientiev, Romanov, & Shishatski 79, also Ramm 86:

$\delta u =$ perturbation in dynamical fields corresponding to perturbation δA in parameters

$$A\frac{\partial \delta u}{\partial t} + P\delta u = -\delta A \frac{\partial u}{\partial t}$$

and

$$\left[A\frac{\partial}{\partial t} + P\right]\left(\frac{\partial u}{\partial t}\right) = \frac{\partial f}{\partial t}$$

Nonlinear Challenges: Why low frequencies are important

Similarly for linearization error - $h > 0$, $u_h =$ fields corresponding to $A + h\delta A$,

$$e = \frac{u_h - u}{h} - \delta u$$

$$A\frac{\partial e}{\partial t} + Pe = -\delta A \frac{\partial}{\partial t}(u_h - u)$$

$$\left[A\frac{\partial}{\partial t} + P\right]\left(\frac{\partial}{\partial t}(u_h - u)\right) = -h\delta A \frac{\partial^2 u_h}{\partial t^2}$$

$$\left[(A + h\delta A)\frac{\partial}{\partial t} + P\right]\left(\frac{\partial^2 u_h}{\partial t^2}\right) = \frac{\partial^2 f}{\partial t^2}$$

RICE

Nonlinear Challenges: Why low frequencies are important

Use causal Green's (inverse) operator:

$$\delta u = -\left[A\frac{\partial}{\partial t} + P\right]^{-1} \delta A \left[A\frac{\partial}{\partial t} + P\right]^{-1} \frac{\partial f}{\partial t}$$

$$e = -h\left[A\frac{\partial}{\partial t} + P\right]^{-1} \delta A \left[A\frac{\partial}{\partial t} + P\right]^{-1} \delta A \left[(A + h\delta A)\frac{\partial}{\partial t} + P\right]^{-1} \frac{\partial^2 f}{\partial t^2}$$

pass to frequency domain:

$$\hat{\delta u} = -[-i\omega A + P]^{-1}\delta A[-i\omega A + P]^{-1} i\omega \hat{f}$$

$$\hat{e} = -h[-i\omega A+P]^{-1}\delta A[-i\omega A+P]^{-1}\delta A[-i\omega A+P]^{-1}(-i\omega)^2\hat{f}+O(h^2\omega^2)$$

RICE

Nonlinear Challenges: Why low frequencies are important

So for small ω,

$$\hat{\delta u} = i\omega P^{-1}\delta A P^{-1}\hat{f} + O(\omega^2)$$

$$\hat{e} = h\omega^2 P^{-1}\delta A P^{-1}\delta A P^{-1}\hat{f} + O(\omega^3)$$

$\hat{f}(0) \neq 0 \Rightarrow$ there exist δA for which

- $P^{-1}\delta A P^{-1}\hat{f} \neq 0$ - δA is *resolved at zero frequency*

$\Rightarrow$ for such δA

- (energy in e) $< O(\|\delta A\|\langle\omega\rangle)$ (energy in δu)

So: linearization error is small $\Rightarrow J_{LS}$ is near-quadratic, for sufficiently low frequency source *and/or* sufficiently small δA.

Further analysis: quadratic directions $\sim$ large-scale features

Linear Challenges: Why reflection is hard

Relative difficulty of reflection vs. transmission

- numerical examples: Gauthier, Virieux & Tarantola 86
- spectral analysis of layered traveltime tomography: Baek & Demanet 11

Spectral analysis of reflection per se: Virieux & Operto 09

Linear Challenges: Why reflection is hard

Reproduction of "Camembert" Example (GVT 86) (thanks: Dong Sun)

Circular high-velocity zone in 1km $\times$ 1km square background - 2% Δv.

Transmission configuration: 8 sources at corners and side midpoints, 400 receivers (100 per side) surround anomaly.

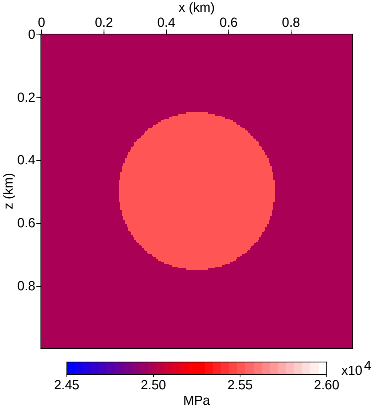
Reflection configuration: all 8 sources, 100 receivers on one side ("top").

Modeling details: 50 Hz Ricker source pulse, density fixed and constant, staggered grid FD modeling, absorbing boundaries.

RICE

Linear Challenges: Why reflection is hard

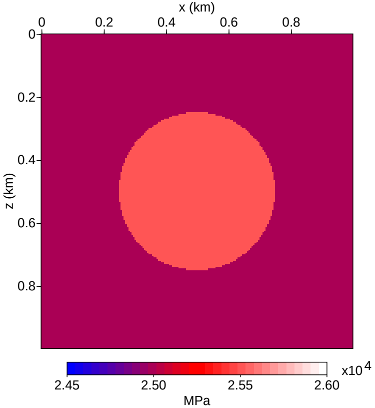
Transmission inversion, 2% anomaly: Initial MS resid $= 2.56\times10^7$; Final after 5 LBFGS steps $= 2.6\times10^5$

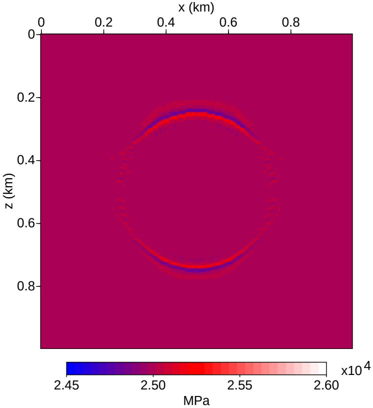


Bulk modulus: Left, model; Right, inverted

Linear Challenges: Why reflection is hard

Reflection configuration: initial MS resid = 3629; final after 5 LBFGS steps = 254





Bulk modulus: Left, model; Right, inverted

Linear Challenges: Why reflection is hard

Message: in reflection case, "the Camembert has melted".

Small anomaly $\Rightarrow$ linear phenomenon

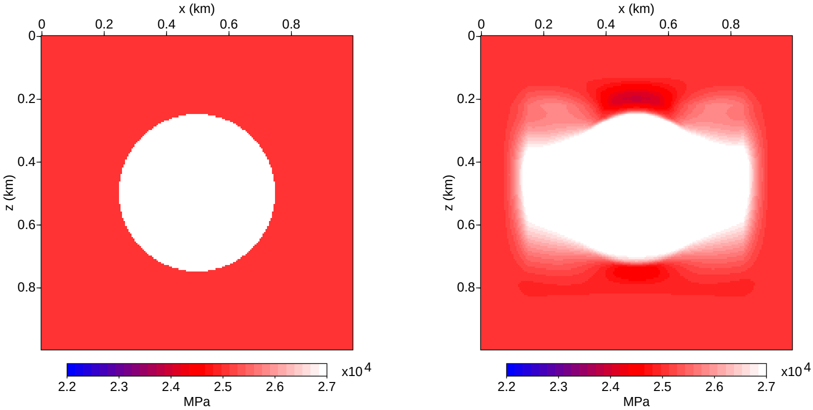
Linear resolution analysis (eg. Virieux & Operto 09): narrow aperture data does not resolve low spatial wavenumbers

Resolution analysis of phase (traveltime tomography) in layered case: Baek & Demanet 11

- model $\mapsto$ traveltime map = composition of (i) increasing rearangement, (ii) invertible algebraic tranformation, (iii) linear operator
- factor (iii) has singular values decaying like $n^{-1/2}$ for diving wave traveltimes, *expontially decaying* for reflected wave traveltimes.

Linear Challenges: Why reflection is hard

Putting it all together: "Large" Camembert (20% anomaly) with 0-60 Hz lowpass filter source. *Continuation in frequency* after Kolb, Collino & Lailly 86 - 5 stages, starting with 0-2 Hz:



Bulk modulus: Left, model; Right, inverted

Agenda

Overview

Challenges for FWI

Extended modeling

Summary

Extended Models and Differential Semblance

Inversion of reflection data: difficulty rel. transmission is *linear* in origin, so look to migration velocity analysis for useful ideas

Prestack migration as approximate inversion: fits subsets of data with non-physical *extended* models (= image volume), so all data matched - no serendipitous local matches!

Transfer info from small to large scales by demanding coherence of extended models

Familiar concept from *depth-domain migration velocity analysis* - independent models (images) grouped together as *image gathers*, coherence $\Rightarrow$ good velocity model

Exploit for automatic model estimation: residual moveout removal (Biondi & Sava 04, Biondi & Zhang 12), van Leeuwen & Mulder 08 (data domain VA), differential waveform inversion (Chauris, poster session), *differential semblance* (image domain VA) S. 86 ...

Extended Models and Differential Semblance

Differential semblance, version 1:

- group data d into gathers $d(s)$ that can be *fit perfectly* (more or less), indexed by $s \in S$ (source posn, offset, slowness,...)
- *extended models* $\bar{M} = \{\bar{m} : S \to M\}$
- *extended modeling* $\bar{F} : \bar{M} \to D$ by

$$\bar{F}[\bar{m}](s) = F[m(s)]$$

- s finely sampled $\Rightarrow$ coherence criterion is $\partial \bar{m}/\partial s = 0$.

The DS objective:

$$J_{DS} = \|\bar{F}[\bar{m}] - d\|^2 + \sigma^2 \left\| \frac{\partial \bar{m}}{\partial s} \right\|^2 + ...$$

RICE

Extended Models and Differential Semblance

Continuation method ($\sigma : 0 \rightarrow \infty$) - theoretical justification Gockenbach, Tapia & S. '95, limits to J_{LS} as $\sigma \rightarrow \infty$.

"Starting" problem: $\sigma \rightarrow 0$, minimizing J_{DS} equivalent to

$$\min_{\bar{m}} \left\| \frac{\partial \bar{m}}{\partial s} \right\|^2 \text{ subj to } \bar{F}[\bar{m}] \simeq d$$

Relation to MVA:

- separate scales: m_0 = macro velocity model (physical), δm = short scale reflectivity model
- linearize: $\bar{m} = m_0 + \delta \bar{m}$, $\bar{F}[\bar{m}] \simeq F[m_0] + D\bar{F}[m_0]\delta \bar{m}$
- approximate inversion of $\delta d = d - F[m_0]$ by migration: $\delta \bar{m} = D\bar{F}[m_0]^{-1}(d - F[m_0]) \simeq D\bar{F}[m_0]^T(d - F[m_0])$

Extended Models and Differential Semblance

$\Rightarrow$ *MVA via optimization:*

$$\min_{m_0} \left\| \frac{\partial}{\partial s} \left[D\bar{F}[m_0]^T (d - F[m_0]) \right] \right\|^2$$

Many implementations with various approximations of $D\bar{F}^T$, choices of s: S. & collaborators early 90's - present, Chauris-Noble 01, Mulder-Plessix 02, de Hoop & collaborators 03-07.

Bottom line: works well when hypotheses are satisfied: linearization (no multiples), scale separation (no salt), *simple kinematics* (no multipathing)

Nonlinear DS with LF control

Drop scale separation, linearization assumptions

Cannot use independent long-scale model as control, as in MVA: "low spatial frequency" not well defined, depends on velocity.

However, *temporal* passband *is* well-defined, and lacks very low frequency energy (0-3, 0-5,... Hz) with good s/n

Generally, inversion is unambiguous if data d is *not* band-limited (good s/n to 0 Hz) - $\bar{F}$ is nearly one-to-one - extended models $\bar{m}$ fitting same data d differ by tradeoff between params, controllable by DS term

So: find a way to supply the low-frequency *data*, as ersatz for long-scale *model* - in fact, *generate* from auxiliary model!

Nonlinear DS with LF control

Define low-frequency source complementary to data passband, *low-frequency (extended) modeling op* F_l ($\bar{F}_l$)

Given *low frequency control model* $m_l \in M$, define extended model $\bar{m} = \bar{m}[d, m_l]$ by minimizing over $\bar{m}$

$$J_{DS}[\bar{m}; d, m_l] = \|\bar{F}[\bar{m}] + \bar{F}_l[\bar{m}] - (d + F_l[m_l])\|^2 + \sigma^2 \left\| \frac{\partial \bar{m}}{\partial s} \right\|^2$$

Determine $m_l \Rightarrow$ minimize

$$J_{LF}[d, m_l] = \left\| \frac{\partial}{\partial s} \bar{m}[d, m_l] \right\|^2$$

(NB: nested optimizations!)

RICE

Nonlinear DS with LF control

$$\min_{m_l} J_{LF}[d, m_l] = \left\| \frac{\partial}{\partial s} \bar{m}[d, m_l] \right\|^2$$

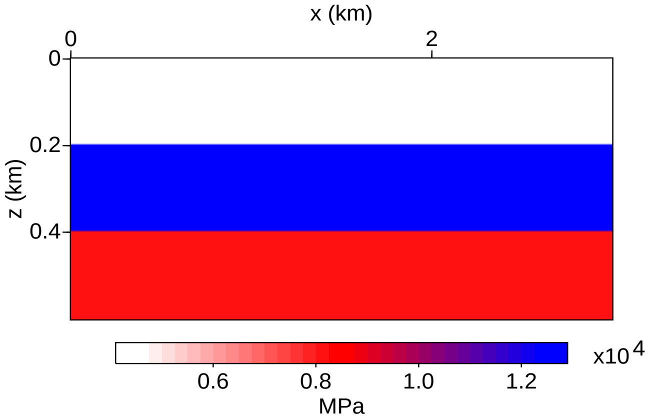
m_l plays same role as migration velocity model, *but no linearization, scale separation assumed*

$\bar{m}[d, m_l]$ analogous to prestack migrated image volume

Initial exploration: Dong Sun PhD thesis, SEG 12, plane wave 2D modeling, simple layered examples, steepest descent with quadratic backtrack.

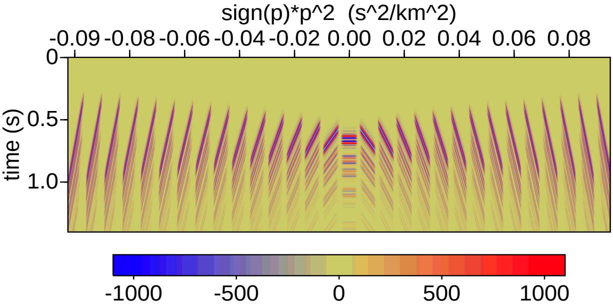
Greatest challenge: efficient and accurate computation of gradient = solution of auxiliary LS problem

Example: DS Inversion with LF control, free surface



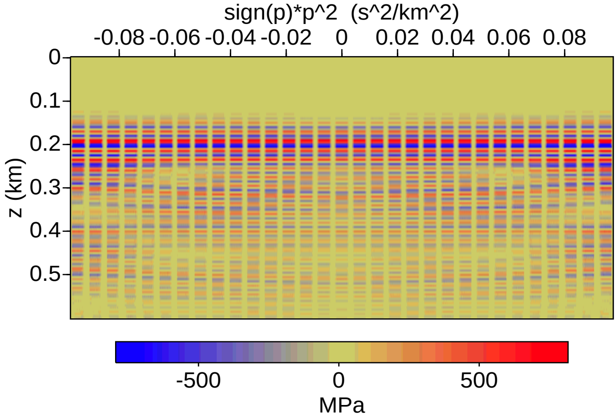
Three layer bulk modulus model. Top surface pressure free, other boundaries absorbing

Example: DS Inversion with LF control, free surface



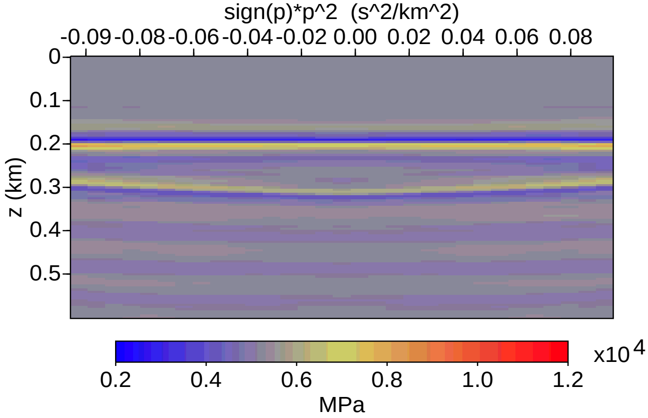
Plane wave data, free surface case

Example: DS Inversion with LF control, free surface



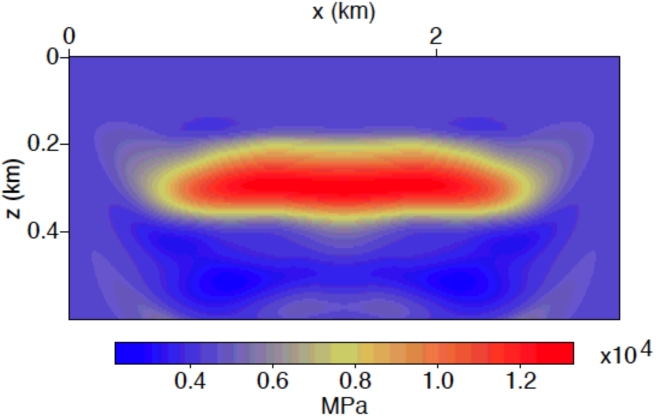
Extended model LS gradient at homog initial model (prestack image volume)

Example: DS Inversion with LF control, free surface



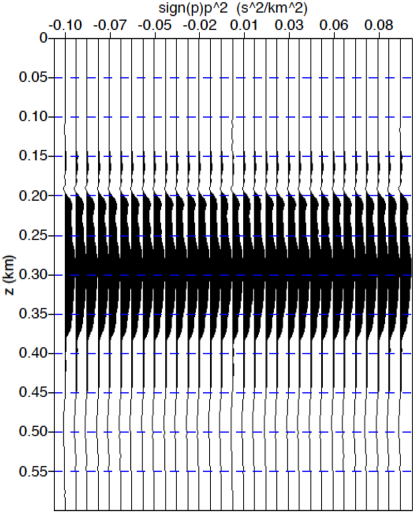
Inverted gather $\bar{m}[d, m_l]$, m_l = homogeneous model, $x = 1.5$ km

Example: DS Inversion with LF control, free surface



Low frequency control model m_l in the 3rd DS-iteration

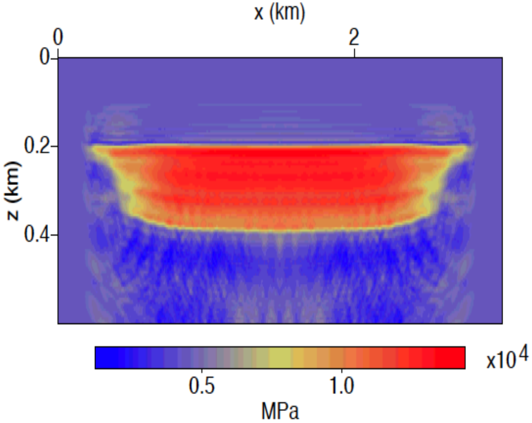
Example: DS Inversion with LF control, free surface



Inverted gather $\bar{m}[d, m_l]$, 3rd DS iteration, $x = 1.5$ km

Example: DS Inversion with LF control, free surface

Standard FWI using stack of optimal DS $\bar{m}$ as initial data
(one-step homotopy $\sigma = 0 \to \infty$)



153 L-BFGS iterations, final RMS error $= 6\%$, final gradient norm < 1 % of original

Space Shift DS

Defect in version 1 of DS already known in MVA context:

> *Image gathers generated from individual surface data bins may not be flat, even when migration velocity is optimally chosen (Nolan & S, 97, Stolk & S 04)*

Source of *kinematic artifacts* obstructing flatness: multiple ray paths connecting sources, receivers with reflection points.

Therefore version 1 of DS only suitable for mild lateral heterogeneity. Must use something else to identify complex refracting structures

Space Shift DS

For MVA, remedy is known: use *space-shift* image gathers $\delta\bar{m}$ (de Hoop, Stolk & S 09)

Claerbout's imaging principle (71): velocity is correct if energy in $\delta\bar{m}(\mathbf{x},\mathbf{h})$ is *focused* at $\mathbf{h}=\mathbf{0}$ ($\mathbf{h}=$ *subsurface offset*)

Quantitative measure of focus: choose $P(\mathbf{h})$ so that $P(\mathbf{0})=0$, $P(\mathbf{h})>0$ if $\mathbf{h}\neq\mathbf{0}$, minimize

$$\sum_{\mathbf{x},\mathbf{h}} |P(\mathbf{h})\delta\bar{m}[m_0](\mathbf{x},\mathbf{h})|^2$$

(e. g. $P(\mathbf{h})=|\mathbf{h}|$).

MVA based on this principle by Shen, Stolk, & S. 03, Shen et al. 05, Albertin 06, 11, Kubir et al. 07, Fei & Williamson 09, 10, Tang & Biondi 11, others - survey in Shen & S 08. Gradient issues: Fei & Williamson 09, Vyas 09.

Space Shift DS

Extension to nonlinear problems - how is $\delta\bar{m}[\mathbf{x},\mathbf{h}]$ the output of an adjoint derivative?

Answer: Replace coefficients m in wave equation with *operators* $\bar{m}$: e. g. $\bar{\kappa}[u](\mathbf{x}) = \int d\mathbf{h}\bar{\kappa}(\mathbf{x},\mathbf{h})u(\mathbf{x}+\mathbf{h})$. *Physical case:* multiplication operators $\bar{\kappa}(\mathbf{x},\mathbf{h}) = \kappa(\mathbf{x})\delta(\mathbf{h})$. Then

$$\delta\bar{m}[m_0] = D\bar{F}[\bar{m}_0]^T(d - F[m])$$

for resulting extended fwd map $\bar{F}$

$\Rightarrow$ Version 2 of nonlinear DS. Physical case = no-action-at-a-distance principle of continuum mechanics = nonlinear version of Claerbout's imaging principle (S, 08). Mathematical foundation: Blazek, Stolk & S. 08.

Agenda

Overview

Challenges for FWI

Extended modeling

Summary

Summary

- restriction to low frequency data makes FWI objective more quadratic, just like you always thought
- transmission inversion is easier than reflection for *linear* reasons, so MVA seems like a good place to look for reflection inversion approaches
- extended modeling provides a formalism for expressing MVA objectives that extend naturally to nonlinear FWI, via *continuation* - provision of starting models, route to FWI solution
- positive early experience with "gather flattening" nonlinear differential semblance
- "survey sinking" NDS involves wave equations with operator coefficients
- Patrick's fingerprints are all over this subject

RICE

Thanks to...

- Florence Delprat and other organizers, EAGE
- my students and postdocs, particularly Dong Sun, Peng Shen, Chris Stolk, Kirk Blazek, Joakim Blanch, Cliff Nolan, Sue Minkoff, Mark Gockenbach, Roelof Versteeg, Michel Kern
- National Science Foundation
- Sponsors of The Rice Inversion Project
- Patrick Lailly, for inspired, inspiring, and fundamental contributions to this field