The Search for a Cycle-Skipping Cure: an Overview

William Symes

Computational and Applied Mathematics & Earth Science, Rice University
The Rice Inversion Project

IPAM, May 2017

Agenda

Fixes within FWI

Alternatives to FWI

Analysis of Extended Waveform Inversion

Prospects

FWI

- Why Low Frequencies are Critical
- Low Frequency Field Recording
- Traveltime Tomography
- Wave Equation Tomography
- Band Extrapolation

FWI - Low Frequencies and Starting Models

Well-established observation, based on heuristic arguments ("cycle skipping"), numerical evidence:

forward modeling operator is *more linear* [FWI objective function is more quadratic] at lower frequencies

Frequency continuation (Kolb, Collino, & Lailly 86): start with lowest frequencies with good S/N

Pratt 99: initial time error $\leq 0.5\lambda$

FWI - Low Frequencies and Starting Models

Constant density acoustics, $m(\mathbf{x})$ = square slowness (compliance), $f(\mathbf{x}, t)$ = finite-energy source, localized at $\mathbf{x}_s$, $u(\mathbf{x}, t; \mathbf{x}_s)$ = acoustic potential for source at $\mathbf{x}_s$,

$$(m\partial_t^2 - \nabla^2)u = f,\ u, f = 0 \text{ for } t < 0$$

pressure $p = \partial_t u$, observed data = $d(\mathbf{x}_r, t; \mathbf{x}_s)$, modeling ("Forward") operator F:

$$F[m](\mathbf{x}_r, t; \mathbf{x}_s) = p(\mathbf{x}_r, t; \mathbf{x}_s)$$

FWI - Low Frequencies and Starting Models

Bare-bones FWI objective = Mean Square Error

$$J_{\mathrm{FWI}}[m,d] = \sum_{\mathbf{x}_s,\mathbf{x}_r} \int dt \, |F[m]-d|^2(\mathbf{x}_r,t;\mathbf{x}_s)$$

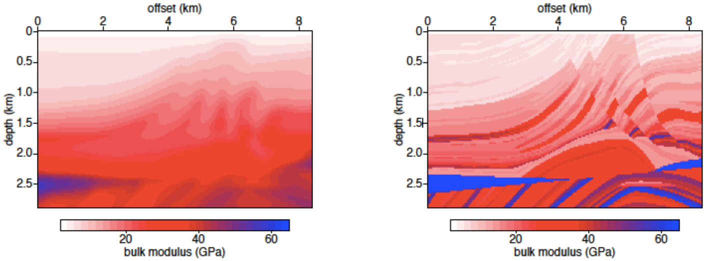
RICE

FWI - Low Frequencies and Starting Models

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

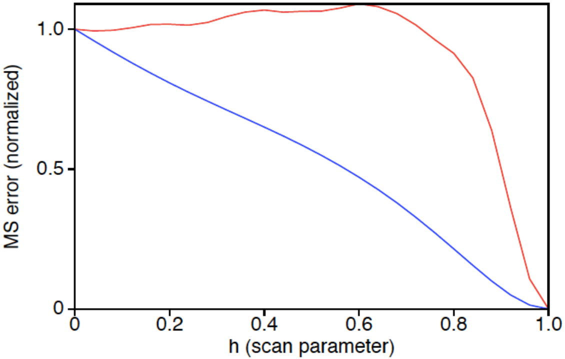
$$f(h) = J_{\mathrm{FWI}}[(1-h)m_0 + hm_1]$$

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



Marmousi bulk modulus: smoothed m_0, original m_1

FWI - Low Frequencies and Starting Models



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

FWI - Low Frequencies and Starting Models

Early hint: Lavrentiev et al 1980 - $[0, \omega_0) \subset$ data passband $\Rightarrow$ data determines model ($=$ const $+$ $C_0^\infty(\mathbf{R}^3)$)

Basic energy estimate, $\| \cdot \| = L^2$ norm:

$$\|\partial_t u(t)\| \le C \|f\|_{[0,t]}$$

($\partial_t u = p =$ pressure, $\frac{1}{2}\|\sqrt{m}\partial_t u\|^2 =$ potential energy)

FWI - Low Frequencies and Starting Models

Linearization, Born approximation, directional derivative field: $\delta u(\mathbf{x}, t; \mathbf{x}_s)$ for model perturbation $\delta m(\mathbf{x})$

$$(m\partial_t^2 - \nabla^2)\delta u = -\delta m \partial_t^2 u, \; \delta u = 0 \text{ for } t < 0$$

more energy estimates:

$$\|\partial_t \delta u\|(t) \le C\|\delta m\|_\infty \|\partial_t^2 u\|_{[0,t]} \le C\|\delta m\|_\infty \|\partial_t f\|_{[0,t]}$$

"linearly good" directions: pick $0 < r < 1$,

$$\|\partial_t \delta u\|(t) \ge rC\|\delta m\|_\infty \|\partial_t f\|_{[0,t]}$$

(NB: $\|\partial_t f\|$ rather than $\|f\|$!)

FWI - Low Frequencies and Starting Models

How good is linear approximation? Yet more energy estimates:

$$\|u[m+\delta m] - u[m] - \delta u\|(t) \leq C\|\delta m\|^2\|\partial_t^2 f\|$$

So ratio *linearization error / linearization* in "linearly good" directions is $\propto \|\delta m\|_\infty \|\partial_t^2 f\|_{[0,t]} / \|\partial_t f\|_{[0,t]}$ - in frequency domain, to $\|\delta m\|_\infty \omega$.

$\Rightarrow$ for directions δm that yield significant δu, FWI objective approaches quadratic as $\|\partial_t^2 f\|_{[0,t]} / \|\partial_t f\|_{[0,t]} \rightarrow 0$, in "linearly good" directions

RICE

FWI - Low Frequencies and Starting Models

Caveat: how do you know that initial model kinematic accuracy is sufficient?

Bunks et al 95: Marmousi variant, homog. initial model $\Rightarrow$ need good S/N at 0.25 Hz

Plessix et al 10: Land survey Mongolia, tomo initial model: success starting at 1.5 Hz, failure starting at 2.0 Hz

Shah 12: visual tests for cycle-skipping

Low Frequency Acquisition

Acquire wavelengths $\geq 2 \times$ travel time error - energetics (Ziolkowski 93)

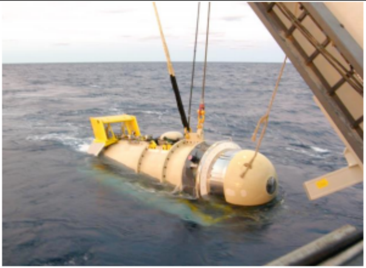


Figure 1: BP's Wolfspar® prototype low-frequency seismic source being retrieved during operational testing at the Gulf of Mexico systems integration test in 2014.

(Thanks: Dellinger et al., SEG 16)

Tomography

obtain large scale features of model from inversion of travel time picks - reflection, diving wave, ...

if residual travel time errors (both for times used in tomography and not!) are $\leq \lambda/2$, then TTT model can initiate FWI

Pratt BP Blind Test EAGE 04 - diving wave TTT (Zelt 97) to construct initial model

Hi-res tomography: CGG Chevron Blind Test SEG 14 - "low" isn't so low any more

RICE

Wave Equation Tomography

Luo & Schuster 91 (...,Kormann et al 16):

$$\Delta\tau(\mathbf{x}_r,\mathbf{x}_s) = \mathrm{argmax}_\tau \int dt\, p(\mathbf{x}_r, t-\tau; \mathbf{x}_s) d(\mathbf{x}_r, t; \mathbf{x}_s)/A(\mathbf{x}_r,\mathbf{x}_s)$$

($A(\mathbf{x}_r,\mathbf{c}_s) =$ max amplitude of $(\mathbf{x}_r,\mathbf{x}_s)$ trace)

$$J_{\mathrm{WET}}[m,d] = \sum_{\mathbf{x}_s,\mathbf{x}_r} |\Delta\tau(\mathbf{x}_r,\mathbf{x}_s)|^2$$

NB traces windowed to single (earliest) arrival $\Rightarrow$ higher likelihood of unique max

$\nabla J_{\mathrm{WET}}[m,d]$ - same adjoint state computation as $\nabla J_{\mathrm{FWI}}[m,d]$, just different input traces (w. $\Delta\tau$ factor)

Band Extrapolation

Li & Demanet 15, 16: FWI with extrapolated low frequency data

identify events via *phase tracking* - fit data with linear comb of "atomic" events, penalize

- strongly for phase(frequency) nonlinearity
- less strongly for phase, dispersion, amplitude, variation with receiver

Having found events, "hang" LF wavelets on them

Warner et al 13 - "frequency down-shifting", identify phase by sparse L1 decon then "hang" LF wavelets

RICE

Agenda

Fixes within FWI

Alternatives to FWI

Analysis of Extended Waveform Inversion

Prospects

Alternatives to FWI

- Laplace-Fourier Domain
- Mass Transport
- Reflection Waveform Inversion
- Reduced Order Models
- Migration Based Travel Time
- Extended Waveform Inversion

Laplace-Fourier domain

Shin et al, 01,...,13:

$$\tilde{p}(\mathbf{x}_r, s, \omega; \mathbf{x}_s) = \int dt\, e^{-i\omega t - st} p(\mathbf{x}_r, t; \mathbf{x}_s)$$

Equiv. to FT of exponentially damped trace - emphasizes early events

Pure Laplace - logarithmic residual: minimize norm of

$$\log\left(\frac{\tilde{p}(\mathbf{x}_r, s, 0; \mathbf{x}_s)}{\tilde{d}(\mathbf{x}_r, s, 0; \mathbf{x}_s)}\right)$$

(wavelet eliminated)

Mass Transport

Metivier, Yang talks

Reflection Waveform Inversion

Mora 89: "Inversion = Migration + Tomography"

Key observation: reflectors in model act as secondary sources, producing transmission response in reflection data

Current FWI iterate includes reflectors ⇒ gradient contains cross-corrs of

- fields propagating in opposite directions - reflection imaging (down/up + up/down)
- fields propagating in same direction - low spatial frequency "tomographic" update (down/down + up/up)

Relection **W**aveform **Inversion** = emphasize 2nd category ("rabbit ears") to update macro-model

RICE

Reflection Waveform Inversion

Wu & Alkhalifah 16: represent source, receiver wave fields in gradient computation as sum of

- up, down, left, right
- finer directional decomposition,
- annular frequency ranges
- scattering angle ranges

Find linear combinations of partial gradient contributions

- minimizing a measure of roughness (eg. biharmonic)
- angle $<< \pi/2$ with standard gradient
- same L2 norm as standard gradient

designer rabbit ears

Reduced Order Models

Mamonov talk

RICE

MBTT

Migration **B**ased **T**ravel **T**ime - Chavent mid-90s, Clement et al 01,..., Tchverda et al 16

confine search in FWI to models of form

$$m = m_0 + \delta m,\ \delta m = W[m_0]DF[m_0]^T r$$

m_0 = smooth background model (updated), r = time reflectivity (updated), $DF[m_0]^T$ = RTM, $W[m_0]$ = preconditioner $\rightarrow$ "true amplitude"

Algorithm: alternating update of m_0, r - model m is by-product

MBTT

Core idea: with true-amplitude migration, will obtain $DF[m_0]\delta m \approx d - F[m_0]$ so at each step solve linearized problem in r update - implicit Gauss-Newton

Effective model $= (m_0, r)$ - much bigger than m!

Modulo linearization error, (could) fit data at every step

$\Rightarrow$ this algorithm really belongs in next section of talk!

Extended FWI - fit everything

Cycle-skipping $\leftrightarrow$ data misfit

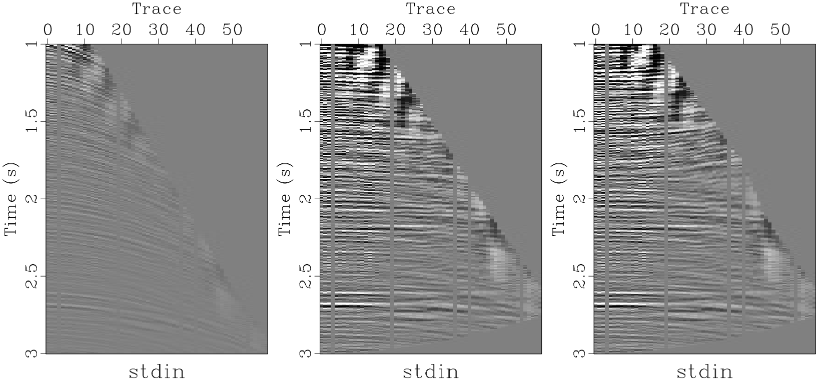
So... add a bunch of parameters, so data is easier to fit - original "physical" model without extra params should be amongst possible *extended models*

"No free lunch" theorem: additional parameters are not in original physics, must be suppressed somehow

This is a very old idea...

Extended FWI - fit everything

CDP 2000 from North Sea 2D survey (Mobil) SEG 94 workshop (Foster & Keys 98)



Left: CDP; Middle: after NMO, "good" v_p; Right: after NMO, "bad" v_p

Extended FWI - fit everything

Convolutional model: after NMO, as function of t_0,

$$d = w * \bar{r} + n$$

presume signature decon has been performed - $w \approx \delta$ - ignore noise

NMO correction is inverse for every trace $\Rightarrow$ data is fit

Each trace is a reflectivity model of the *same* earth midpoint - so all should be same ("physical")

$\Rightarrow$ "gathers flat" - pick your favorite flatness criterion

RICE

Extended FWI - fit everything

Common formalism for extended modeling:

m = physical model

$\bar{m}$ = extended model

$\bar{F}$ = extended modeling operator

A = "annihilator" - null space = physical models

E = "extension map" - turns physical model into extended model

RICE

Extended FWI - fit everything

For convolutional model,

$m = (v_p, r)$ - single traces for each midpoint, used to model all offsets

$F[m] =$ inverse NMO(v_p) applied to midpoint r trace generates all offsets

$\bar{m} = (v_p, \bar{r})$ - independent $\bar{r}$ trace for every midpoint, offset

$\bar{F}[\bar{m}] =$ inverse NMO(v_p) applied to each $\bar{r}$ trace for every midpoint, offset

$E[m]$: repeat same r trace for every offset in $\bar{r}$

$A[\bar{m}] = \partial_h \bar{r}$ (for example)

RICE

Extended FWI Catalog (incomplete!!!)

(blue = talk in PM):

- surface: model acquisition gathers independently - all models same (gathers flat)
  - shot record (Kern & S. 94)
  - common offset (Mulder & ten Kroode 01, Chauris & Noble 02)
- subsurface: model param becomes operator (action at distance) - physical models diagonal (gathers focused)
  - space lag (Stolk & de Hoop 01, Shen & S. 08)
  - time lag (Biondi & Almomin 14)
- source: add params to source functions
  - source-receiver: AWI (Warner & Guasch 14,16, Huang & S. 16)
  - space-time: WRI (Herrmann & van Leeuwen 13, Wang & Yingst 16)
  - volume, surface: Huang & S. 15,16

RICE

Agenda

Fixes within FWI

Alternatives to FWI

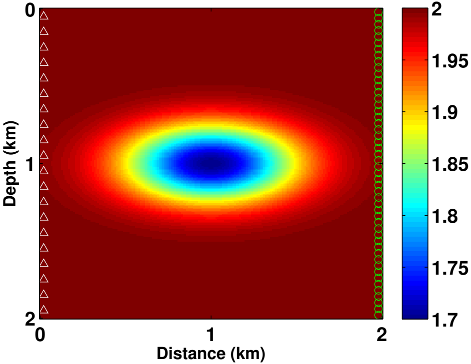
Analysis of Extended Waveform Inversion

Prospects

RICE

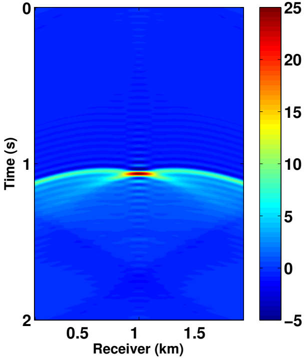
A Simple Example

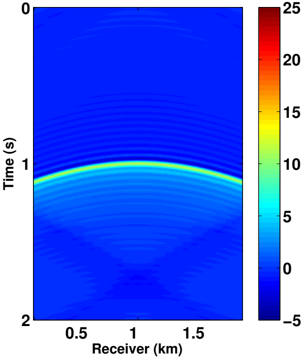
Transmission: sources left, receivers right.



$m = v$ field, f=source (3-20 Hz bandpass), $F[m] = G[v] * f$

A Simple Example

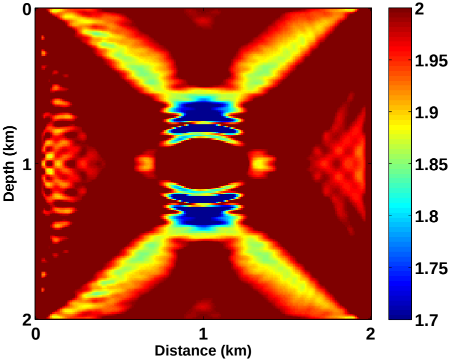




Left: 1 km shot gather, Right: same in initial model (homog.)

A Simple Example

Result of FWI, 21 LBFGS iterations



A Simple Example

Source-receiver extension: $\bar{m}=(v,\ \bar{f}(\mathbf{x}_r,t;\mathbf{x}_s))$ = independent source function for each trace)

$\bar{F}[\bar{m}](\mathbf{x}_r,t;\mathbf{x}_s) = (G[v_p] * \bar{f})(\mathbf{x}_r,t;\mathbf{x}_s)$

Single-arrival case: $G(v_p)(\mathbf{x}_r,t;\mathbf{x}_s) \approx a(\mathbf{x}_r,\mathbf{x}_s)\delta(t-\tau(\mathbf{x}_r,\mathbf{x}_s)$ - high-frequency approx.

$\bar{F}[\bar{m}] = a\bar{f}(t-\tau) + O(\lambda) = \bar{S}[v]\bar{f}$ - linear ($\bar{f}$) and nonlinear (v) parameters

Assumption on "physical" source f: concentrated near $t=0$ - so $\|tf\|^2 \approx 0 \Rightarrow$ use $A =$ multiply $\bar{f}$ by t

A Simple Example

Aim: $\bar{S}[v]\bar{f} \approx d$ and $A\bar{f} \approx 0$

Nested approach (essential! see Y. Huang thesis 2016):

Define $\bar{f}_\alpha[v, d] =$ reg. LS minimizer of

$$\|\bar{S}[v]\bar{f} - d\|^2 + \alpha^2\|\bar{f}\|^2$$

$$\bar{f}_\alpha[v] = N[v]_\alpha^{-1}\bar{S}[v]^T d$$

$$N[v]_\alpha = \bar{S}[v]^T\bar{S}[v] + \alpha^2 I$$

A Simple Example

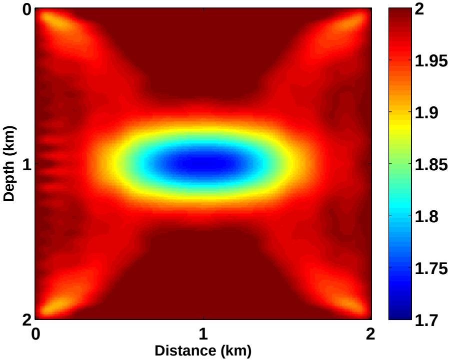
Extended FWI objective $J_\alpha[v,d] = \frac{1}{2}\|A\bar{f}_\alpha[v,d]\|^2$

Estimate v by minimizing $J_\alpha[v,d]$ - get $\bar{f}_\alpha[v,d]$ as by-product, estimate physical f by stacking

Why is this any better than FWI?

A Simple Example

Result of Extended FWI, 21 LBFGS iterations



Analysis of Gradient

Gradient requires second least squares solution:

$$\bar{g}_\alpha[v,d] = N[v]_\alpha^{-1} A^T A \bar{f}_\alpha[v,d]$$

$$\nabla_v J_\alpha[v,d] = -DS[v]^T(\bar{f}_\alpha[v,d], \bar{S}[v]\bar{g}_\alpha[v,d])$$
$$+DS[v]^T(\bar{g}_\alpha[v,d], d - \bar{S}[v]\bar{f}_\alpha[v,d])$$

$DS[v]^T(\bar{f}, d) =$ RTM operator with source $\bar{f}$ acting on data d

(what a mess...)

RICE

Analysis of Gradient

How to see what this means...

Temporarily ignore amplitude, redefine $\bar{S}[v]$ as

$$\bar{S}[v]\bar{f} = \bar{f}(t-\tau) + O(\lambda)$$

$\Rightarrow$ $\bar{S}[v]$ is essentially unitary:

$$\bar{S}[v]^T\bar{S}[v] = I + O(\lambda)$$

Note: could not be true of S!

Analysis of Gradient

$\bar{S}$ essentially unitary $\Rightarrow$ factorization lemma

$$D\bar{S}[v]\delta v = \bar{S}[v](Q[v]\delta v) + O(\lambda)$$

- $Q[v] + Q[v]^T = O(\lambda)$ (ess. skew-adjoint)
- $Q[v]$ is of order 1, that is, acts like $O(1/\lambda)$ in freq domain

For source-receiver extension, relation with kinematics:

$$Q[v]\delta v = -(D\tau[v]\delta v)\frac{\partial}{\partial t}$$

RICE

Analysis of Gradient

Use factorization and $\bar{S}^T\bar{S} \approx I$ - most of gradient collapses, leaving

$$\langle \nabla_v J_\alpha[v,d], \delta v \rangle = \langle \bar{S}[v]^T d, [A^T A, Q[v]\delta v]\bar{S}[v]^T d \rangle + O(\lambda)$$

Low-noise data: $d = S[v^*]f = f(t - \tau[v^*]) + O(\lambda)$

$=$ ... (Song & S. TRIP 94 for details) $=$ ...

$$= \|f\|^2 \sum_{\mathbf{x}_r,\mathbf{x}_s} (\tau[v] - \tau[v^*]) D\tau[v]\delta v + O(\lambda)$$

RICE

Analysis of Gradient

$$= \|f\|^2 D_v J_{\mathrm{TT}}[v, \tau[v^*]]\delta v + O(\lambda)$$

where

$$J_{\mathrm{TT}}[v, \tau^{\mathrm{obs}}] = \frac{1}{2}\sum_{\mathbf{x}_r, \mathbf{x}_s}(\tau[v] - \tau^{\mathrm{obs}})^2$$

is the usual travel time tomography objective. In other words,

$$\nabla_v J_\alpha[v, d] \propto \nabla_v J_{\mathrm{TT}}[v, \tau[v^*]] + O(\lambda)$$

Minimize $J_\alpha \Rightarrow$ *tomography without picking*

RICE

Analysis of Gradient

A couple of dangling ends...

Can't actually ignore amplitude - consequence of including it is $Q + Q^T = O(1)$ (not $O(\lambda)$)

For source-receiver extension, unitary property and $Q + Q^T = O(\lambda)$ recovered via AWI objective (Warner & Guasch 14, 16):

$$J_{\texttt{AWI}}[v,d] = \frac{1}{2}\sum_{\mathbf{x}_r,\mathbf{x}_s} \frac{\|A\bar{f}_\alpha[v,d](\mathbf{x}_r,\cdot;\mathbf{x}_s)\|^2}{\|\bar{f}_\alpha[v,d](\mathbf{x}_r,\cdot;\mathbf{x}_s)\|^2}$$

up to $O(\lambda)$ error, same as amplitude-free case $\Rightarrow$ $\nabla J_{\mathrm{AWI}} \propto \nabla J_{\mathrm{TT}}$

RICE

Analysis of Gradient

Also single arrival assumption extremely restrictive, *but* multiple arrivals destroy

- essentially unitary property
- factorization lemma
- kinematic content of $D\bar{S}$

$\Rightarrow$ source-receiver ext'd FWI suffers FWI-like failure to converge with energetic multiple arrivals (G. Huang & S. 17)

Analysis of Gradient

Fix: different extension - give up point source model, spread source energy over surface, penalize spread *in space*, recover *all properties* listed above even in presence of caustics (G. Huang & S., SEG 16) (also no need to assume that source known)

Analysis is *generic*: applies to model extensions as well, eg. subsurface offset for reflection data (ten Kroode 12, S. 12).

Hou & S 15, 16: unitary formulation of Born-based subsurface offset extended FWI

Agenda

Fixes within FWI

Alternatives to FWI

Analysis of Extended Waveform Inversion

Prospects

RICE

Outer gradient for nested formulations

Recall: energy in $p \propto$ energy in source, but energy in $\delta p \propto$ energy in ∂_t source

If inner (linear) problem must be solved by an iteration based on S, convergence typically not in sense of continuity of DS

Examples of negative effect on outer (v) gradient computation - theses of Y. Huang, E. Cocher.

Stable gradient approx for unitary S - S. TRIP 16

Reflection via Source Extension

Extended FWI based on source extensions - successful inversion of reflection data when FWI fails (AWI - Warner & Guasch 14, 16; surface source extension, WRI - Herrnann & van Leeuwin 13, Wang & Yingst 16; surface source extension - G. Huang & S. 16).

Why?

[So far, theory only for transmission]

Hint: surface source extension for non-smooth media, unitary extended modeling op via Dirichlet-to-Neumann map, time reversal, energy decay. Construction of ess. skew Q - how to relate $[A^T A, Q]$ to kinematics? What kinematics?

An embarrassment of riches

Is cycle-skipping a solved problem?

Some benchmarking - thanks to IFP, SEG/EAGE, Chevron & many others

However, little *comparative* benchmarking - none *reproducible* (by you!)

Is this possible? Desirable? Requires open source? Algorithms sufficiently well-defined that they can be implemented to operate independently of their "parents"?

Cf. Madagascar migration gallery - is an inversion gallery feasible?

Acknowledgements

- Guanghui Huang
- TRIP members & sponsors
- WS 2 Org Comm
- IPAM
- You!