Extended Waveform Inversion

William W. Symes

The Rice Inversion Project
Computational and Applied Mathematics
Rice University

WS 9 - EAGE 2017
Extended modeling permits data to be fit well throughout inversion process

Penalty for model extension should $\rightarrow 0$ as inversion progresses, so weight should $\rightarrow \infty$

Discrepancy principle - keep data misfit near nominal data noise level - adjusts penalty weight, implies convergence extended WI $\rightarrow$ full WI
Collaborators

Lei Fu - PhD 2016
A discrepancy based penalty method for extended waveform inversion: Lei Fu, William W. Symes, GEOPHYSICS, Posted online on 25 May 2017


Jie Hou - PhD 2016
Agenda

Waveform Inversion in Born Approximation

Extended Waveform Inversion

The Discrepancy Principle and Convergence to FWI

Perspectives
Born Approximation

\[ m = \text{background model} \]

\[ \delta m = \text{model perturbation} \]

\[ F[m] = \text{Born modeling operator} \]

\[ F[m] \delta m = \text{Born data from } m, \delta m \]
**Born Approximation**

Example: Acoustic Born Approximation

\[ m = v(x)^2 \text{ (background squared velocity), } w(t) = \text{ source wavelet}, \ p(x, t; y) = \text{ pressure field = solution of point radiator problem} \]

\[
\left( \frac{\partial^2}{\partial t^2} - v(x)^2 \nabla^2 \right) p(x, t; y) = w(t) \delta(x - y)
\]

\[ S = \text{ sampling operator } x = x_r \text{ survey receiver locations } y = x_s \text{ source locations} \]
Born Approximation

\[ \delta m = r(x) \text{ reflectivity} = \text{squared velocity perturbation} \]
\[ \delta(v(x)^2) \]

\[ \delta p(x, t; y) = \text{pressure perturbation} \]

\[ \left( \frac{\partial^2}{\partial t^2} - v^2 \nabla^2 \right) \delta p = r \nabla^2 p \]

\[ F[m] \delta m = S \delta p \]
Born FWI Inversion Example

Adapted from SEG/EAGE overthrust model. Background velocity $v(x)$:
Born FWI Inversion Example

Reflectivity = velocity perturbation $r(x)$:
## Born FWI Inversion Example

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source wavelet</td>
<td>bandpass $5 - 20$ Hz</td>
</tr>
<tr>
<td>Source position $x_s$</td>
<td>$x : 1 - 7$ km @ 40 m, $z = 40$ m</td>
</tr>
<tr>
<td>Receiver position $x_r$</td>
<td>$x : 0 - 8$ km @ 40 m, $z = 0$ m</td>
</tr>
<tr>
<td>Space and time</td>
<td>$x = 8$ km, $z = 2$ km, $t = 3$ s</td>
</tr>
<tr>
<td>Grid</td>
<td>$dx = dz = 20$ m</td>
</tr>
<tr>
<td>Time step</td>
<td>$dt = 2$ ms</td>
</tr>
<tr>
<td>Initial velocity</td>
<td>$v = 1.5$ km/s</td>
</tr>
</tbody>
</table>
Born FWI Inversion Example

Shot gather $d_{75}$ ($x_s = 4$ km)
Born FWI Inversion Example

Born FWI:

Nested approach to avoid mismatched sensitivities

\[ \delta m[m] = \text{minimizer over } \delta m \text{ of } \| F[m] \delta m - d \|_2^2 \text{ (least squares migration)} \text{ - 20 CG iterations} \]

\[ m = \text{minimizer over } m \text{ of } \| F[m] \delta m[v] - d \|_2^2 \text{ - 20 steepest descent iterations} \]
Born FWI Inversion Example

Born FWI - initial velocity $v_0$
Born FWI Inversion Example

Born FWI - velocity after 20 steepest descent steps
Born FWI Inversion Example

Born FWI - reflectivity = least squares migration (20 CG its) after 20 steepest descent steps
Agenda

Waveform Inversion in Born Approximation

Extended Waveform Inversion

The Discrepancy Principle and Convergence to FWI

Perspectives
Subsurface offset extension

Concept:

cycle-skipping $\rightarrow$ poor data fit in LSM unless $\nu$ is nearly correct

so add parameters to model so data can always be fit

Subsurface offset extension - model perturbation becomes operator ("action at distance")


Subsurface offset extension

\[ \delta \tilde{m} = \text{extended model perturbation} \]

\[ \tilde{F}[m] = \text{extended Born modeling operator} \]

\[ \tilde{F}[m] \delta \tilde{m} = \text{Born data from } m, \delta \tilde{m} \]

Extension operator \( E : \delta m \mapsto \delta \tilde{m} \)

Extension property: \( \tilde{F}[m] E \delta m = F[m] \delta m \)
Subsurface offset extension

Subsurface offset extension: $\delta m = \bar{r}(x, h)$, $h =$ subsurface offset

Green’s function representation ($h = (h, 0)$):

$$\bar{F}[m] \delta m(x_r, t; x_s) =$$

$$\int dx \int dh \int d\tau G(x_r, t - \tau; x + h) r(x, h) \nabla^2_x G(x_s, \tau; x - h)$$

Stock & de Hoop 01: adjoint of survey-sinking migration (Claerbout, 85), RTM version
Subsurface offset extension

Data Fitting Property: for “reasonable” $d$, $m$, possible to find $\delta \bar{m}$ so that

$$\bar{F}[m] \delta \bar{m} \approx d$$
Subsurface offset extension

For $m = \nu_0(x)$, use this $\delta \bar{m} = \bar{r}(x, h)\ldots$
Subsurface offset extension

...and get this predicted data $\tilde{F}[m] \delta \bar{m}$, residual $\tilde{F}[m] \delta \bar{m} - d$ w/ 20 CG iterations:
Have added parameters to model \((h)\), have to get rid of them somehow.

Extension operator: \( Er(x, h) = r(x)\delta(h) \)

⇒ physical (non-extended) modeling acts only at \(h = 0\) - focused

Penalty for failure to focus: \(A\delta\tilde{m} = h\tilde{r}(x, h)\)
Extended Inversion by Variable Projection

Extended Waveform Inversion: minimize over $m, \delta \bar{m}$

$$\| \tilde{F}[m] \delta \bar{m} - d \|^2 + \alpha \| A \delta \bar{m} \|^2$$

Variable projection (Golub & Pereyra 03, Mulder & van Leeuwen 10): nested minimization

- inner minimization over $\delta \bar{m}$ (extended LSM) to obtain $\delta \bar{m}_\alpha[m]$
- outer minimization over $m$ of

$$J_{EWI}[m] = \| \tilde{F}[m] \delta \bar{m}_\alpha[m] - d \|^2 + \alpha \| A \delta \bar{m}_\alpha[m] \|^2$$

$$= e(\alpha) + \alpha p(\alpha)$$
Extended Inversion by Variable Projection

Variable projection method - pluses, minuses:

+: simple gradient formula, assuming exact minimization for $\delta \bar{m}_\alpha[m]$:

$$\nabla J_{EWI}[m] = 2D\bar{F}[m]^*(\delta \bar{m}_\alpha[m], \bar{F}[m]\delta \bar{m}_\alpha[m] - d)$$

$D\bar{F}[m]^* = \text{“tomographic operator” (Biondi-Sava 04)}$ - RTM-like computation

-: must choose $\alpha$ somehow
Agenda

Waveform Inversion in Born Approximation

Extended Waveform Inversion

The Discrepancy Principle and Convergence to FWI

Perspectives
The Discrepancy Principle in practice

Concept: Choose nominal error level $X$, bounds $X_- < X < X_+$. Demand $X_- < e(\alpha) < X_+$

As $m$ updates, eventually $e(\alpha) < X_-$

Update formula: a bit of calculus shows that

$$\alpha_+ = \alpha_c + \frac{X_+ - e(\alpha_c)}{2p(\alpha_c)}$$

assures $\alpha_+ > \alpha_c$ and $e(\alpha_+) < X_+$ - “safe” $\alpha$ update

Can use this for $\alpha = 0$ - know can achieve $e(0) < X_+$ - self-starting
The Discrepancy Principle in practice

Initial $m$

- Calculate $e(\alpha = 0)$
- Estimate $X_-$ and $X_+$

Update $\alpha_+ = \alpha + \frac{x_+ - e(\alpha)}{2p(\alpha)}$

- Calculate $e(\alpha_+)$

$e \in [X_-, X_+]$ True

$e(\alpha) > X_-$ True

- Update $m$

$\alpha_+ \neq 2$ or $\alpha_+ \neq 1.5$ False

- Calculate $e(\alpha)$

$e(\alpha) = 0$ Estimate $!!$ and $!!$

Initial $!!$

Update $!! = ! + !!(!)$

Calculate $!!(!)$

$!! \in [!!!, !!!!!]$ Update $!!$ True

False

$!! \ast = 2$ or $!!/ = 1.5$

Calculate $!!(!)$

$!!(!) > !!$ True

False
The Discrepancy Principle in practice

![Graph showing relative data misfit and iteration number for different values of α.]

- $\alpha = 0$
- $\alpha = 1.1 \times 10^{-6}$
- $\alpha = 3.2 \times 10^{-5}$
- $\alpha = 1.1 \times 10^{-4}$
- $\alpha = 2.4 \times 10^{-4}$

$X^+ = 6.9\%$

$X^- = 4.0\%$
The Discrepancy Principle in practice

![Graph showing the Discrepancy Principle]

- $\alpha = 1.1 \times 10^{-6}$
- $\alpha = 1.1 \times 10^{-6}$
- $\alpha = 3.2 \times 10^{-5}$
- $\alpha = 1.1 \times 10^{-4}$
- $\alpha = 2.4 \times 10^{-4}$
The Discrepancy Principle in practice

Estimated velocity - 20 steepest descent steps
The Discrepancy Principle in practice

Target velocity
The Discrepancy Principle in practice

Inverted Reflectivity ($\delta \tilde{m}_\alpha[m]$) = ext’d LSM at 20 s.d. steps
The Discrepancy Principle in practice

Estimated Reflectivity for initial velocity $\delta \tilde{m}_\alpha[m], m = v_0(x)$
The Discrepancy Principle in practice

Strict test of kinematic validity; perform *non-ext’d* LSM at inverted velocity model

Then re-simulate data - does it match (with no extension to help fit!)???
The Discrepancy Principle in practice

Non-extended LSM at inverted velocity
The Discrepancy Principle in practice

Target Reflectivity

Distance (km)

Depth (km)

Velocity (km/s)
The Discrepancy Principle in practice

Comparison of three traces from data (black) and re-simulation from non-extended LSM at inverted velocity (red)
Agenda

Waveform Inversion in Born Approximation

Extended Waveform Inversion

The Discrepancy Principle and Convergence to FWI

Perspectives
Why does it work?

In $\alpha \to 0$ limit ("IVA"), subsurface offset EWI asymptotically $\approx$ variant of stereotomography (S., ten Kroode IPTA 14), with gradient - same optimization problem

$J_{EWI} = $ data-weighted mean square error in arrival time, slowness

Continuation $\alpha = 0 \to \infty$ - Gockenbach et al. 95

Need better theoretical grounding for (stereo)tomography - does it exhibit local mins? (apparently not)
What about cost?

Example used 20 outer iterations, each requiring 20 inner iterations

= 400 extended modelings/migrations of entire data (not counting line-search evaluations)

Forget it...
What about cost?

Not so fast:

- ν updates → increased focus → shrink $h - axis$ while maintaining discrepancy criterion
- shorter $h$ axis $\sim$ larger $\alpha$ $\Rightarrow$ stable condition of LS problem
- early updates with low frequency band (3 - 7.5 Hz), subsampling in space/time, increase frequency in stages (3 - 15 Hz, 3 - 30 Hz) and refine grid as model converges

Upshot: 2 orders of magnitude speedup for typical 2D problem (Fu & S. Geophysics 17)
What about cost?

Furthermore

- accelerate inner iteration via approximate inverse (Hou & S. Geophysics 16)
- steepest descent inefficient - use closer relative to Newton, preconditioner

Inner loop - 10x speedup

Outer loop - projected 10x speedup
How do you find $X$?

Proposal: use the discrepancy principle

Either $p(\alpha) \to 0$ with $X_- < e(\alpha) < X_+ \Rightarrow X$ is correct noise estimate

Or reducing $p(\alpha)$ requires $e(\alpha) > X_+ \Rightarrow$ increase $X$

Or reducing $p(\alpha)$ requires $e(\alpha) < X_- \Rightarrow$ decrease $X$

$\Rightarrow$ discover noise level in course of algorithm
Scope of algorithm?

Chief requirements:

- good data fit by solving inner problem - start with $\alpha = 0$
- drive penalty term to 0: value at optimum is known

These are features of all extended WI algorithms - shot record, plane wave, time-shift (Biondi & Almomin), AWI (Warner & Guasch), WRI (Herrmann & van Leeuwen, Wang & Yingst),...
Summary

Extended modeling permits data to be fit well throughout inversion process

Penalty for model extension should $\rightarrow 0$ as inversion progresses, so weight should $\rightarrow \infty$

Discrepancy principle - keep data misfit near nominal data noise level - adjusts penalty weight, implies convergence extended WI $\rightarrow$ full WI
Thanks to...

- Lei Fu, Jie Hou, TRIP students and postdocs
- Shell for supporting Jie’s PhD studies
- Current and former sponsors of TRIP
- Organizers of WS 09
- Guy Chavent