Wavefield-Reconstruction Inversion - WRI

Felix J. Herrmann
Strategy

Derive an alternative *extended* formulation:

- *fits* data for *poor* starting models
- less prone to local minima
- computationally feasible
- relaxes the physics while staying solidly grounded
Equation error approach

If we “know” the wavefields everywhere, we could solve for $\mathbf{m}$ from

$$A(\mathbf{m}) \mathbf{u}_i = \mathbf{q}_i$$

via

$$\min_{\mathbf{m}} \| A(\mathbf{m}) P_i^{-1} \mathbf{d}_i - \mathbf{q}_i \|_2^2 \quad \left( \text{cf.} \min_{\mathbf{m}} \| P_i A(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i \|_2^2 \right)$$

which is linear in $\mathbf{m}$.

The challenge is to reconstruct wavefields from partial measurements...

[Richter, ’81]
wave-equation \times \text{wavefield} = \text{source}

\text{versus}

\left(\text{wave-equation} \times \text{wavefield}\right) \times \text{wavefield} = \left(\text{source} \times \text{data}\right)
observed data  initial data  data-augmented solution
WRI – Wavefield-Reconstruction Inversion

For $\mathbf{m}$ fixed, reconstruct wavefields by jointly fitting observed shots

$$ P\mathbf{u}_i \approx \mathbf{d}_i $$

and wave-equations

$$ A(\mathbf{m})\mathbf{u}_i \approx \mathbf{q}_i $$

via least-squares solutions of the data-augmented wave-equation

$$ \min_{\mathbf{u}_i} \left\| \begin{pmatrix} P_i \\ A(\mathbf{m}) \end{pmatrix} \mathbf{u}_i - \begin{pmatrix} \mathbf{d}_i \\ \mathbf{q}_i \end{pmatrix} \right\|_2^2 $$

followed by fixing $\mathbf{u}_i$ and solving

$$ \min_{\mathbf{m}} \left\| A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i \right\|_2^2 $$

[van Leeuwen & FJH, 2013]
PDE-constrained optimization
all-at-once full-space approach

\[
\min_{m,u} \sum_{i=1}^{M} \| P_i u_i - d_i \|^2_2 \quad \text{s.t.} \quad A_i(m)u_i = q_i
\]

- avoids having to solve the PDE explicitly
- sparse (GN) Hessian
- requires storing all variables \((m,u)\)
- does not scale to industry-scale seismic problems
Adjoint-state/reduced-space formulation

Elimination of the constraint leads for all sources to

$$\min_{\mathbf{m}} \phi_{\text{red}}(\mathbf{m}) = \sum_{i=1}^{M} \| P_i A_i(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i \|_2^2$$

- no need to store all wavefields (block-elimination)
- suitable for black-box optimization (e.g., l-BFGS)
- need to solve forward & adjoint PDEs
- very non-linear dependence on earth model ($\mathbf{m}$)
- dense (GN) Hessian, involves additional PDE solves
- reliance on accurate starting models to avoid cycle skipping
Instead of eliminating, we add constraints as penalties—i.e.,

\[
\min_{m,u} \phi_\lambda(m, u) = \sum_{i=1}^{M} \|Pu_i - d_i\|_2^2 + \lambda^2 \|A_i(m)u_i - q_i\|_2^2
\]

coincides with original problem when \( \lambda \uparrow \infty \)

- no need to store all the fields \((u)\)
- no adjoint solves
- sparse approximation of Gauss-Newton Hessian for small \(\lambda\)
- less non-linear in \(m\)
- need to solve data-augmented wave equation
Related work

Contrast-source formulation
- combined objective is similar
- but does not eliminate wavefields via variable projection [Golub ’03, van Leeuwen & Aravkin ’12]
- requires storage of wavefields for all sources

Tomographic extensions
- sensitivities to “noise” & relative strengths of events
- WRI uses wave equation itself to “focus”
Variable projection

Solve data-augmented wave equation for each source

\[
\begin{pmatrix}
P_i \\
\lambda A_i(m)
\end{pmatrix}
\begin{pmatrix}
u_i, \lambda \\
\lambda q_i
\end{pmatrix}
\approx
\begin{pmatrix}
d_i \\
\lambda q_i
\end{pmatrix}
\]

Define reduced objective with proxy wavefields

\[
\phi_\lambda(m) = \phi_\lambda(m, \bar{u}_\lambda) = \|P\bar{u}_\lambda - d\|_2^2 + \lambda^2 \|A(m)\bar{u}_\lambda - q\|_2^2
\]

[Aravkin & van Leeuwen, ’12; van Leeuwen & FJH, ’13]
WRI method

for each source $i$

solve $\begin{pmatrix} P_i \\ \lambda A_i(m) \end{pmatrix} u_{\lambda,i} \approx \begin{pmatrix} d_i \\ \lambda q_i \end{pmatrix}$

$g = g + \lambda^2 \omega^2 \text{diag}(\bar{u}_{i,\lambda})^*(A(m)\bar{u}_{i,\lambda} - q_i)$

end

$m = m - \alpha g$

Conventional method

for each source $i$

solve $A(m)u_i = q_i$

solve $A(m)^*v_i = P_i^*(P_iu_i - d_i)$

$g = g + \omega^2 \text{diag}(u_i)^*v_i$

end

$m = m - \alpha g$
Bas Peters, Felix J. Herrmann, and Tristan van Leeuwen, “Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion”

WRI method

for each source $i$

\[
\text{solve } \left( \begin{array}{c} P_i \\ \lambda A_i(m) \end{array} \right) u_{\lambda,i} \approx \left( \begin{array}{c} d_i \\ \lambda q_i \end{array} \right)
\]

\[
g = g + \lambda^2 \omega^2 \text{diag}(\bar{u}_i,\lambda)^*(A(m)\bar{u}_i,\lambda - q_i)
\]

\[
H_{GN} = H_{GN} + \lambda^2 \omega^4 \text{diag}(u_i)^*\text{diag}(u_i)
\]

end

\[
m = m - \alpha H^{-1}_{GN}g
\]

diagonal Hessian
pseudo Hessian

Conventional method

for each source $i$

\[
\text{solve } A(m)u_i = q_i
\]

\[
\text{solve } A(m)^*v_i = P_i^*(P_iu_i - d_i)
\]

\[
g = g + \omega^2 \text{diag}(u_i)^*v_i
\]

end

\[
m = m - \alpha g
\]
dense Hessian
& too expensive
One reflector example

true model
Wavefields in *homogeneous* background

**FWI**
- Forward wavefield $\bar{u}$
- Adjoint wavefield $\bar{v}$
- Reconstructed wavefield $\bar{u}_\lambda$

**WRI**
- PDE residual $\bar{v}_\lambda$
Wavefields in *homogeneous* background

**FWI**
- **forward**
- **adjoint**

**WRI**
- **reconstructed wavefield**
- **PDE residual**

\[ \bar{u}, \bar{v}, \bar{u}_\lambda, \bar{v}_\lambda \]
Wavefields in *homogeneous* background
Wavefields in *homogeneous* background

**FWI**
- forward: $\tilde{u}$
- adjoint: $\bar{v}$

**WRI**
- reconstructed wavefield: $\tilde{u}_\lambda$
- PDE residual: $\bar{v}_\lambda$
Wavefields in *homogeneous* background

**FWI**
- forward
- adjoint
- reconstructed wavefield

**WRI**
- PDE residual
Observations

Objective of fitting *both* data & wave equation
- introduces (reflection) events in our wavefield reconstructions
- we use these events to update the velocity model with the wave equation

Corresponds to a variable-projection approach solving for $u$ & $m$, respectively.

Differs from reduced/adjoint formulation where
- these events are absent
- velocity is highly nonlinear in the data
Local minima

single shot, single frequency data for linear velocity profile $v(z) = v_0 + \alpha z$,

misfit as function of $(v_0, \alpha)$
FWI vs WRI

Solutions of data-augmented system force data fits... **no longer cycle skipped!**
Extended modelling

The penalty formulation

$$\min_{\tilde{m},u} ||P\mathbf{u} - \mathbf{d}||^2_2 + \lambda^2 ||A(m)\mathbf{u} - \mathbf{q}||^2_2$$

can be interpreted as

$$\min_{\tilde{m}} \text{misfit}(\tilde{m}) + \text{annihilator}(\tilde{m})$$

with

$$\tilde{m} = (m, u)$$

For a physically plausible model we have

$$\text{annihilator}(\tilde{m}) = 0$$

[Symes, personal communication]
Examples
– FWI is known to fail

Velocity models with
  ‣ low-velocity “kick backs”
  ‣ high-contrast high-velocity unconformities

Solve WRI w/ poor starting models using
  ‣ multiple frequency sweeps w/ warm starts
  ‣ additional convex constraints

Use WRI to leverage reflected energy during the inversions...
BG Compass model

- **Challenges:** velocity kick backs & detailed geology
- Isotropic acoustic data & poor starting model
- Invert for slownesses w/ acoustic kernel
- 24 frequency batches {5 6}, {6 7}, ..., {28 29} Hz w/ 5 frequencies each
- 103 sources/receivers w/ 55m sample interval
- l-BFGS with 15 iterations per frequency band
- Two frequency sweeps
True & initial model
FWI vs WRI

Result FWI

Result WRI, $\lambda = 1$
Gradients

First update FWI

First update WRI, $\lambda = 1$
First sweep
Second sweep
Objective function value

Objective WRI, cycle 1

Objective WRI, cycle 2

Data fit increases at some iterates
WRI vs. FWI

Larger # of degrees of freedom

“more convex”
Chevron blind test data

**Inversion strategy:**
1. Frequency domain WRI with Source estimation;
3. Batch sizes of random frequency subsets: 3, 6, 10, 10;
4. Batch size of random source subsets: 300;
5. Optimization solver: l-BFGS with 30 iterations per frequency band;
6. 2 passes of WRI from frequency 3-11 Hz;
7. Grid size: 20m;
8. Minimum offset used: 1000m;
9. No pre-processing !!!

[van Leeuwen & FJH, ’13; ’15, Peters et. al. ’14]
Chevron blind test data

Zhilong Fang  Xiang Li  Bas Peters  Brendan Smithyman  Mengmeng Yang  Felix J. Herrmann
Initial model
Initial data fit
— @3 Hz & shot 800

\[ \lambda = 1e3 \]
Model update

Depth [km]

Lateral [km]

-400
-200
0
200
400

-400
-200
0
200
400

0
5
10
15
20
25
30
35
40
45

0
1
2
3
4
5

0
5
10
15
20
25
30
35
40
45
Source wavelet comparison

- **Amplitude**
  - Frequency [Hz]: 0, 5, 10, 15
  - True Wavelet
  - Estimated Wavelet

- **Phase**
  - Frequency [Hz]: 0, 5, 10, 15
  - True Wavelet
  - Estimated Wavelet
Kirchhoff migration
—Initial model
Kirchhoff migration
—Inversion result
WRI w/ density & curvelet sparsity

Depth [km]

Lateral [km]

Vt, Vini, Vfinal
Model update
Observations

WRI obtains a reasonable inversion results for the velocity & source

No data preprocessing or extensive handholding needed

*How does this method hold up in cases where we have high-contrast high-velocity unconformities such as salt?*
BP benchmark

- **Challenge**: high-contrast & high-velocity unconformity
- *Added convex constraints*
- Isotropic acoustic “inverse-crime” data w/ known 15 Hz Ricker wavelet
- Invert for slownesses w/ acoustic kernel
- Frequency bands: 3–20Hz in overlapping batches of 2
- Number of sources: 126; number of receivers: 299
- Maximum number of outer iterations per frequency batch: 25
- Maximum number of inner iterations for convex subproblems: 2000
- 8 frequency sweeps that relax the constraints
WRI
w/ or w/o convex constraints
Conclusions

New method for wave-equation based inversion:

- benefits from same extended search space as in *all-at-once* but w/ memory & CPU requirements of adjoint-state approaches
- fits the observed data by design & is “less non-linear”
- therefore less susceptible to local minima

Experiments show that WRI succeeds where FWI fails because it uses

- reflected energy to invert for low-velocity kickbacks
- convex constraints by virtue of the (near) diagonal GN Hessian

Candidate for “automatic” salt flooding...
This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, DownUnder GeoSolutions, Hess, Petrobras, PGS, Subsalt Ltd, WesternGeco, and Woodside.

Thank you for your attention!

https://www.slim.eos.ubc.ca/