A dual formulation for time-domain wavefield reconstruction inversion

Gabrio Rizzuti\textsuperscript{*,1}, Mathias Louboutin\textsuperscript{1}, Rongrong Wang\textsuperscript{2}, Emmanouil Daskalakis\textsuperscript{3}, and Felix J. Herrmann\textsuperscript{1}

\textsuperscript{1} Georgia Institute of Technology
\textsuperscript{2} Michigan State University
\textsuperscript{3} University of British Columbia
PDE-constrained optimization

\[
\inf_{m,u} \frac{1}{2} \| d - R u \|^2 \quad \text{s.t.} \quad A(m)u = S^* f
\]

\[
\begin{align*}
d & : \text{data} & A(m) & : \text{wave equation} \\
f & : \text{source} & R & : \text{receiver interp.} \\
m & : \text{model} & S & : \text{source interp.} \\
u & : \text{wavefield}
\end{align*}
\]
PDE-constrained optimization (all-at-once full-space)

\[ \mathcal{L}(m, u, v) = \frac{1}{2} \| d - Ru \|^2 + \langle v, S^* f - A(m) u \rangle \]

[Haber, E., and Ascher, U. M., 2001; Biros, G., and Ghattas, O., 2005; Grote et al, 2011]

Pros:
- **simultaneous** update of model, wavefield, multiplier
- **cheap** evaluation and gradient computation
- sparse Hessian

Cons:
- **storage of fields and multipliers**
  (for each source and time/frequency sample)
- **updates** computationally demanding
PDE-constrained optimization (reduced space)

\[
J(m) = \frac{1}{2} \left\| d - F(m) S^* f \right\|^2 \\
F(m) = RA(m)^{-1}
\]

forward map


Pros:
- low(er) **storage** requirements

Cons:
- highly **non-convex** (needs good starting model)
- requires exact **PDE solutions**
- dense Hessian
PDE-constrained optimization (penalty method)

\[ J_\lambda(m, u) = \frac{1}{2} \| d - Ru \|^2 + \frac{\lambda^2}{2} \| S^* f - A(m) u \|^2 \]

PDE penalty

[van den Berg, P. M., and Kleinman, R. E., 1997; van Leeuwen, T. and Herrmann, F. J., 2013]
PDE-constrained optimization (wavefield reconstruction inversion)

\[
\bar{J}_\lambda(m) = \frac{1}{2} \| d - R \bar{u} \|^2 + \frac{\lambda^2}{2} \| S^* f - A(m) \bar{u} \|^2
\]

\[
\begin{bmatrix}
R \\
\lambda A(m)
\end{bmatrix} \bar{u} = \begin{bmatrix}
d \\
\lambda S^* f
\end{bmatrix}
\]

augmented wave equation

[van Leeuwen, T. and Herrmann, F. J., 2013]

Pros:
✓ no simultaneous field storage for all sources
✓ sparse Gauss-Newton Hessian

Cons:
✗ continuation strategy on \( \lambda \)
✗ need augmented PDE solution:
  frequency domain: proper preconditioning?
  time domain: no explicit time-marching scheme
Question: feasible model extension?

- convenient handling of slack variables
- feasible computation of augmented PDE (e.g., can be solved by time-marching)
Wavefield Reconstruction Inversion (time domain approximation)

\[ J_\lambda(m) = \frac{1}{2} \| d - F(m) \bar{q} \|^2 + \frac{\lambda^2}{2} \| S^* f - \bar{q} \|^2 \]

\[
\begin{bmatrix}
F(m) \\
\lambda I
\end{bmatrix}
\bar{q} =
\begin{bmatrix}
d \\
\lambda S^* f
\end{bmatrix}
\]

augmented wave equation

[Wang et al, 2016]

\( \bar{q} \sim \) generalized source
(see also [Huang et al, 2018])

First order Taylor expansion wrt \( 1/\lambda^2 \):

\[ \bar{q} \approx S^* f + \frac{1}{\lambda^2} F(m)^* (d - F(m) S^* f), \quad \lambda \to +\infty \]
Denoising reformulation of WRI

$$\inf_{m,u} \frac{1}{2} \left\| S^* f - A(m)u \right\|^2 \quad \text{s.t.} \quad \left\| d - Ru \right\| \leq \varepsilon$$

PDE misfit

data constraint

[Wang, R., and Herrmann, F. J., 2017]
Dual formulation of WRI - Lagrangian

\[
\inf_{m,u} \frac{1}{2} \left\| S^* f - A(m)u \right\|^2 \quad \text{s.t.} \quad \left\| d - Ru \right\| \leq \varepsilon
\]

PDE misfit \hspace{2cm} \text{data constraint}

**Saddle-point** problem

\[
\inf_{m,u} \sup_{y} \mathcal{L}(m, u, y)
\]

\[
\mathcal{L}(m, u, y) = \frac{1}{2} \left\| S^* f - A(m)u \right\|^2 + \langle y, d - Ru \rangle - \varepsilon \| y \|
\]
Dual formulation of WRI – Augmented wave equation

Solving for $u$ ...

$$A(m)\overline{u} = S^* f + F(m)^* y$$

augmented wave equation

Data interfere with the system!

Saddle-point problem

$$\inf_{m,u} \sup_y \mathcal{L}(m, u, y) \geq \inf_{m} \sup_y \inf_u \mathcal{L}(m, u, y)$$

$$\mathcal{L}(m, u, y) = \frac{1}{2} \| S^* f - A(m)u \|^2 + \langle y, d - R u \rangle - \varepsilon \| y \|$$
Transmission data problem:

- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- $y =$ background data residual
Dual formulation of WRI – Augmented wave equation

Transmission data problem:

- 1 km x 1 km model
- Background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- y = background data residual
Transmission data problem:

- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- y = background data residual
Dual formulation of WRI – Augmented wave equation

**Transmission data** problem:

- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- $y =$ background data residual
Dual formulation of WRI – Objective and gradients

Dual saddle-point formulation

\[
\inf_m \sup_y \bar{\mathcal{L}}(m, y) = \frac{1}{2} \|F(m)^* y\|^2 + \langle y, d - F(m) S^* f \rangle - \varepsilon \|y\|
\]

Gradients

\[
\nabla_m \bar{\mathcal{L}} = -\text{Jac}(m, S^* f + F(m)^* y) y
\]
\[
\nabla_y \bar{\mathcal{L}} = d - F(m)(S^* f + F(m)^* y)
\]

\[
\text{Jac}(m, q) = \left. \frac{d}{dm} (\cdot \mapsto F(\cdot) q) \right|_m
\]

~ conventional FWI gradient

generalized-source data residual
Dual formulation of WRI (recap)

Dual saddle-point formulation

\[
\inf_m \sup_y \bar{\mathcal{L}}(m, y) \\
\bar{\mathcal{L}}(m, y) = -\frac{1}{2} \| F(m)^* y \|^2 + \langle y, d - F(m) S^* f \rangle - \varepsilon \| y \|
\]

Obtained model extension along data space:

\[\bar{\mathcal{L}} : M \times D \rightarrow \mathbb{R}\]

- amenable to time-domain methods
- extra variable storage is affordable
- no continuation strategy needed for relaxation parameter

- extra time complexity (2x PDE solutions, wrt FWI)
- dense Hessian
- need to figure out best optimization strategy (work in progress ...)

(*)
Dual formulation of WRI – Dual variable scaling

Dual saddle-point formulation

\[
\inf_m \sup_y \bar{\mathcal{L}}(m, y) = -\frac{1}{2} \| F(m)^* y \|^2 + \langle y, d - F(m) S^* f \rangle - \varepsilon \| y \|.
\]

Preconditioning/\textbf{Scaling issue} in Lagrangian/augmented wave equation:

\[
A(m) \bar{u} = S^* f + \alpha F(m)^* y
\]

Adaptive strategy?
Dual formulation of WRI – Dual variable scaling

Dual saddle-point formulation

\[
\inf_{m} \sup_{y} \bar{\mathcal{L}}(m, y) = -\frac{1}{2} \|F(m)^*y\|^2 + \langle y, d - F(m)S^*f \rangle - \varepsilon \|y\| \\
\bar{\mathcal{L}}(m, y) = \langle y, r \rangle - \varepsilon \text{sign}(|\langle y, r \rangle|) \|y\|,
\]

\[
\|F(m)^*y\|^2 > \varepsilon \|y\|, \quad |\langle y, r \rangle| \geq \varepsilon \|y\|
\]

r: data residual

Solving for the scaling parameter

\[
\tilde{\mathcal{L}}(m, y, \alpha) := \bar{\mathcal{L}}(m, \alpha y)
\]

\[
\bar{\mathcal{L}}(m, y) := \sup_{\alpha} \tilde{\mathcal{L}}(m, y, \alpha) = \tilde{\mathcal{L}}(m, y, \alpha(m, y))
\]

\[
\alpha(m, y) = \left\{ \begin{array}{ll}
\langle y, r \rangle - \varepsilon \text{sign}(\langle y, r \rangle) \|y\| & , \quad |\langle y, r \rangle| \geq \varepsilon \|y\| \\
\frac{\langle y, r \rangle - \varepsilon \text{sign}(\langle y, r \rangle) \|y\|}{\|F(m)^*y\|^2} & , \quad \|F(m)^*y\|^2 > \varepsilon \|y\| \\
0, & , \quad \text{otherwise}
\end{array} \right.
\]
Dual formulation of WRI (scaled)

\[
\inf_m \sup_y \bar{\mathcal{L}}(m, y)
\]

\[
\bar{\mathcal{L}}(m, y) = \begin{cases} 
\frac{1}{2} (|\langle \hat{y}, r \rangle| - \varepsilon \|\hat{y}\|)^2, & |\langle \hat{y}, r \rangle| \geq \varepsilon \|\hat{y}\| \\
0, & \text{otherwise}
\end{cases}
\]

\[
(\hat{y} = \frac{y}{\|F(m)^* y\|})
\]
**Diving-wave** example:

- 5 km x 10 km model
- linear velocity, varying slope:
  
  \[
  v(x, z) = v_0 + \beta z \\
  v_0 = 2 \text{ km/s}, \quad 0.5 \leq \beta \leq 1
  \]
- single source (5 Hz peak), 21 receivers
  
  - Plot misfit landscape as a function of slope
Diving-wave example:

- 5 km x 10 km model
- linear velocity, varying slope
- single source (5 Hz peak), 21 receivers

Plot misfit landscape as a function of slope
Diving-wave example:

- 5 km x 10 km model
- linear velocity
- single source (5 Hz peak), 21 receivers

Plot misfit landscape as a function of slope

Spurious minima
Diving-wave example:

- 5 km x 10 km model
- linear velocity, varying slope
- single source (5 Hz peak), 21 receivers

Plot misfit landscape as a function of slope

Spurious minima
Reconstruction algorithm (sketch)

Alternating gradient descent

Given \( m_0, y_0 = d - F(m_0)S^*f \)

for \( n = 1: \text{Niter} \)

\[
\begin{align*}
\text{compute gradient wrt } m: & \quad \nabla_{m_n} \bar{\mathcal{L}} \\
\text{update } m \text{ (minimize):} & \quad m_n \leftarrow m_n - \beta \nabla_{m_n} \bar{\mathcal{L}} \\
\text{compute gradient wrt } y: & \quad \nabla_{y_n} \bar{\mathcal{L}} \\
\text{update } y \text{ (maximize):} & \quad y_n \leftarrow y_n + \gamma \nabla_{y_n} \bar{\mathcal{L}}
\end{align*}
\]
BG Compass (reflections/diving waves) [preliminary results]

**BG compass:**

- ~ 3 km x 6 km model
- 25 sources (10 Hz peak), 251 receivers
- *challenging* for conventional FWI!

**Optimization setup:**

- *gradient descent* (fixed step length)
- just 10 iterations
BG Compass  [preliminary results]
BG Compass (worse background) [preliminary results]
Presented a reconstruction algorithm potentially apt to large 3D problems:

- based on model extension along data dimension: robust to spurious minima
- storage of slack variables is affordable (unlike full-space methods)
- computationally advantageous:
  PDEs can be solved in time domain with explicit schemes (unlike classical WRI)

To do:

- study of optimization strategy
- implement constraints


van Leeuwen, T., and Herrmann, F. J., Mitigating local minima in full-waveform inversion by expanding the search space, Geophysical Journal International 195.1 (2013)

Peters, B., and Herrmann, F. J., and van Leeuwen, T., Wave-equation Based Inversion with the Penalty Method-Adjoint-state Versus Wavefield-Reconstruction Inversion, 76th EAGE Conference (2014)

