A dual formulation for time-domain wavefield reconstruction inversion

Gabrio Rizzuti*,1, Mathias Louboutin1, Rongrong Wang2, Emmanouil Daskalakis3, and Felix J. Herrmann1

1 Georgia Institute of Technology
2 Michigan State University
3 Vancouver Community College
PDE-constrained optimization

\[
\begin{align*}
\min_{m,u} & \quad \frac{1}{2} \| d - R u \|^2 \\
\text{subject to} & \quad A(m) u = q
\end{align*}
\]

Vectors:
\[
\begin{align*}
d & \quad \text{data} \\
q & \quad \text{source} \\
m & \quad \text{model} \\
u & \quad \text{wavefield}
\end{align*}
\]

Operators:
\[
\begin{align*}
A(m) & \quad \text{wave equation} \\
R & \quad \text{receiver restriction}
\end{align*}
\]
PDE-constrained optimization (all-at-once full-space)

\[
\max_v \min_{\mathbf{m}, \mathbf{u}} \mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{v}), \quad \mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{v}) = \frac{1}{2} \| \mathbf{d} - R \mathbf{u} \|^2 + \langle \mathbf{v}, \mathbf{q} - A(\mathbf{m}) \mathbf{u} \rangle
\]

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

Pros:
✓ cheap evaluation and gradient computation (no need for PDE solution)

Cons:
✗ need simultaneous storage of wavefields and multipliers (for each source and time/frequency sample)
PDE-constrained optimization (reduced space)

$$\min_{\mathbf{m}} J(\mathbf{m}), \quad J(\mathbf{m}) = \frac{1}{2} \| \mathbf{d} - F(\mathbf{m}) \mathbf{q} \|^2, \quad F(\mathbf{m}) = R A(\mathbf{m})^{-1}$$


Pros:
✓ no simultaneous wavefield storage for all sources and frequencies (compute and discard)

Cons:
✗ highly non-convex (needs good starting model)
✗ requires exact PDE solutions (prohibitive in 3D for frequency domain)
PDE-constrained optimization (penalty method)

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

PDE penalty

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

\[ J_\lambda(m) = \frac{1}{2} \| d - R \tilde{u} \|^2 + \frac{\lambda^2}{2} \| f - A(m) \tilde{u} \|^2 \]

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

Pros:
✓ no simultaneous **wavefield storage** for all sources and frequencies
   (compute and discard)

Cons:
✗ needs **augmented PDE solution**
   **frequency** domain: does not effectively scale to 3D
   **time** domain: no explicit time-marching scheme
Early attempt to circumvent WRI shortcomings

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

[Wang et al, 2016, Huang et al, 2018]:

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

augmented wave equation

Variable projection approximation:

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

\[ \min_{m, u} \frac{1}{2} \|q - A(m) u\|^2 \quad \text{s.t.} \quad \|d - R u\| \leq \varepsilon \]

PDE misfit \hspace{10cm} data constraint

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

\[
\min_{m,u} \frac{1}{2} \|q - A(m) u\|^2 \quad \text{s.t.} \quad \|d - R u\| \leq \varepsilon
\]

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

Saddle-point problem:

\[
\max_y \min_{m,u} \mathcal{L}(m, u, y) = \frac{1}{2} \|q - A(m) u\|^2 + \langle y, d - R u \rangle - \varepsilon \|y\|
\]
Dual formulation of WRI – Augmented wave equation

Solving for $\mathbf{u}$ ...

$$A(\mathbf{m}) \tilde{\mathbf{u}} = \mathbf{q} + F(\mathbf{m})^* \mathbf{y}$$

augmented wave equation

Saddle-point problem:

$$\max_{\mathbf{y}} \min_{\mathbf{m}, \mathbf{u}} \mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{y}) = \max_{\mathbf{y}} \min_{\mathbf{m}} \mathcal{L}(\mathbf{m}, \tilde{\mathbf{u}}, \mathbf{y})$$

$$\mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{y}) = \frac{1}{2} \| \mathbf{q} - A(\mathbf{m}) \mathbf{u} \|^2 + \langle \mathbf{y}, \mathbf{d} - R \mathbf{u} \rangle - \varepsilon \| \mathbf{y} \|$$
Dual formulation of WRI – Objective and gradients

Dual saddle-point formulation (= after $\mathbf{u}$ substitution):

$$\max_y \min_m \bar{\mathcal{L}}(m, y)$$

$$\bar{\mathcal{L}}(m, y) = -\frac{1}{2} \|F(m)^* y\|^2 + \langle y, d - F(m) q \rangle - \varepsilon \|y\|$$

Gradients:

$$\nabla_m \bar{\mathcal{L}} = -D F(m, q + F(m)^* y)^* y$$

similar to conventional FWI gradient

$$\nabla_y \bar{\mathcal{L}} = d - F(m) (q + F(m)^* y) - \varepsilon y / \|y\|$$

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

Dual saddle-point formulation:

$$\max_y \min_m \tilde{L}(m, y)$$

$$\tilde{L}(m, y) = -\frac{1}{2} ||F(m)^* y||^2 + \langle y, d - F(m) q \rangle - \varepsilon ||y||$$

Obtained **model extension** along **data space**: $\tilde{L} : M \times D \rightarrow \mathbb{R}$

Pros:

- amenable to **time-domain** methods
- extra variable **storage is affordable**

Cons:

- extra **time complexity** (2x PDE solutions wrt FWI)
- non trivial **optimization strategy**
Other issues (1): Dual variable scaling

Dual saddle-point formulation:

$$\max_y \min_m \bar{L}(m, y)$$

$$\bar{L}(m, y) = -\frac{1}{2} \|F(m)^* y\|^2 + \langle y, d - F(m) q \rangle - \varepsilon \|y\|$$

$$A(m) \bar{u} = q + \alpha F(m)^* y$$

unbalanced contributions of physical and augmented sources
Other issues (1): Dual variable scaling

Dual saddle-point formulation:

\[
\max_y \min_m \tilde{\mathcal{L}}(m, y) \quad \text{s.t. } \quad \tilde{\mathcal{L}}(m, y) = -\frac{1}{2} \| F(m)^* y \|^2 + \langle y, d - F(m) q \rangle - \varepsilon \| y \| 
\]

Solving for the scaling parameter...

\[
\tilde{\mathcal{L}}(m, y, \alpha) := \tilde{\mathcal{L}}(m, \alpha y) \\
\tilde{\mathcal{L}}(m, y) := \max_\alpha \tilde{\mathcal{L}}(m, y, \alpha) = \tilde{\mathcal{L}}(m, y, \bar{\alpha}) \quad \text{(variable projection!)}
\]

\[
\alpha = \begin{cases} 
\text{sign}(\langle y, r \rangle) \frac{\| y \| - \varepsilon \| y \|}{\| F(m)^* y \|^2}, & \langle y, r \rangle \geq \varepsilon \| y \| \\
0, & \text{otherwise}
\end{cases}
\]

r: data residual
Other issues (1): Dual variable scaling

Dual saddle-point formulation (scaled):

\[
\max_m \min_y \bar{\mathcal{L}}(m, y)
\]

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

\[
y = \frac{y}{\|F(m)^* y\|}
\]

Gradients:

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

\[
\nabla_y \bar{\mathcal{L}} = \alpha \nabla_y \bar{\mathcal{L}}(m, \alpha y)
\]
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

- **alternating update** approach: ineffective
- **variable projection for y** (fixed m) = WRI: expensive

\[
F(m) F(m)^* \bar{y} + \varepsilon \bar{y} / ||\bar{y}|| = r(m)
\]
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

- **alternating update** approach: ineffective
- **variable projection for y** (fixed m) = WRI: expensive

\[
F(m) F(m)^* \bar{y} + \varepsilon \bar{y} / ||\bar{y}|| = r(m)
\]

\[
\bar{L}_\varepsilon(m, y) = -\frac{1}{2} ||F(m)^* y||^2 + \langle y, r(m) \rangle - \varepsilon ||y||
\]

\[
\bar{L}_\lambda(m, y) = -\frac{1}{2} ||F(m)^* y||^2 + \langle y, r(m) \rangle - \frac{\lambda^2}{2} ||y||^2
\]

To avoid non-linear system in y
Theoretical/numerical studies evidence:

\[ \mathbf{y} \approx \mathbf{r} \text{ data residual cheap approximation of the optimal } \mathbf{y} \]

Reduced formulation:

\[
\min_{\mathbf{m}} \bar{\mathcal{L}}(\mathbf{m})
\]

\[
\bar{\mathcal{L}}(\mathbf{m}) = \bar{\mathcal{L}}(\mathbf{m}, \mathbf{r}(\mathbf{m})) = \frac{1}{2} \frac{||\mathbf{r}||^2}{||F(\mathbf{m})^* \mathbf{r}||^2} \left( ||\mathbf{r}|| - \varepsilon \right)^2
\]

Gradient:

\[
\nabla_{\mathbf{m}} \bar{\mathcal{L}} = \nabla_{\mathbf{m}} \bar{\mathcal{L}}(\mathbf{m}, \mathbf{r}) - D F(\mathbf{m}, \mathbf{q})^* \nabla_{\mathbf{y}} \bar{\mathcal{L}}(\mathbf{m}, \mathbf{r})
\]

See also [van Leeuwen, 2019]
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

✓ $y \approx r$ data residual cheap approximation of the optimal $y$
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

✓ $y \approx r$ data residual cheap approximation of the optimal $y$
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

✓ \( y \approx r \) data residual cheap approximation of the optimal \( y \)
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

✓ \( y \approx r \) data residual cheap approximation of the optimal \( y \)
Other issues (2): Optimization strategy

Theoretical/numerical studies evidence:

✓ \( y \approx r \) data residual cheap approximation of the optimal \( y \)
Other issues (3): Weighted PDE misfit

Yet another important theme!

prior information about source position: \textbf{weighted PDE misfit} ([Huang et al, 2018]):

$$\min_{m, u} \frac{1}{2} \| q - A(m) u \|_W^2 \quad \text{s.t.} \quad \| d - R u \| \leq \varepsilon$$

$$\| u \|_W^2 := \langle W^{-1} u, u \rangle, \quad W^{-1} = \text{diag}(w^{-1}), \quad w^{-1}(x; x_s) = \frac{\| x - x_s \|^2 + \delta^2}{\delta^2}$$
Other issues (3): Weighted PDE misfit

Yet another important theme!

- prior information about source position: **weighted PDE misfit** ([Huang et al, 2018]):

\[
\bar{L}_W(m, y) = -\frac{1}{2} \| F(m)^* y \|_{W^{-1}}^2 + \langle y, r(m) \rangle - \varepsilon \| y \|
\]

\[
\| u \|_{W^{-1}}^2 := \langle W u, u \rangle, \quad W = \text{diag}(w), \quad w(x; x_s) = \frac{\delta^2}{\| x - x_s \|^2 + \delta^2}
\]
Other issues (3): Weighted PDE misfit

Yet another important theme!

- prior information about source position: weighted PDE misfit ([Huang et al, 2018]):

\[
\min_{m,u} \|q - A(m) u\|_{1,W} \quad \text{s.t.} \quad \|d - R u\| \leq \varepsilon \quad \text{(alternatively)}
\]

\[
\|u\|_{1,W} := \|W^{-1/2} u\|_1
\]
Other issues (3): Weighted PDE misfit

Yet another important theme!

- prior information about source position: \textbf{weighted PDE misfit} ([Huang et al, 2018]):

\[
\min_{\mathbf{m}, \mathbf{u}} \| \mathbf{q} - A(\mathbf{m}) \mathbf{u} \|_{1,W} + \frac{1}{2\mu} \| \mathbf{q} - A(\mathbf{m}) \mathbf{u} \|_{2,W}^2 \quad \text{s.t.} \quad \| \mathbf{d} - R \mathbf{u} \| \leq \varepsilon
\]

\[
\| \mathbf{u} \|_{1,W} := \| W^{-1/2} \mathbf{u} \|_1
\]

similarly to [Sharan et al, 2019]
Numerical examples

Caveat:

inversion experiments carried out in the **frequency domain:**
- computational convenience
- fair comparison with conventional WRI (only feasible in frequency domain)
Numerical examples – Gaussian lens [Huang et al, 2018]

Source/receiver configuration: 50 sources (top), 200 receivers (bottom)
Optimization strategy: Single frequency (6 Hz, wavelength ~ 333 m), Algorithm: L-BFGS (20 iters)
Numerical examples – Gaussian lens [Huang et al, 2018]

Source/receiver configuration: 50 sources (top), 200 receivers (bottom)
Optimization strategy: Single frequency (6 Hz, wavelength ~ 333 m), Algorithm: L-BFGS (20 iters)
Numerical examples – Gaussian lens [Huang et al, 2018]

Source/receiver configuration: 50 sources (top), 200 receivers (bottom)
Optimization strategy: Single frequency (6 Hz, wavelength ~ 333 m), Algorithm: L-BFGS (20 iters)
Source/receiver configuration: 50 sources, ~300 receivers
Optimization strategy: Multiscale, frequency range: 5 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Numerical examples – BG Compass model

Source/receiver configuration: 50 sources, ~300 receivers
Optimization strategy: Multiscale, frequency range: 5 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Numerical examples – BG Compass model

Source/receiver configuration: 50 sources, ~ 300 receivers
Optimization strategy: Multiscale, frequency range: 5 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Numerical examples – BG Compass model

Source/receiver configuration: 50 sources, ~ 300 receivers
Optimization strategy: Multiscale, frequency range: 5 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Numerical examples – Marmousi model

Source/receiver configuration: 100 sources, ~ 850 receivers

Optimization strategy: Multiscale, frequency range: 3 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Numerical examples – Marmousi model

Source/receiver configuration: 100 sources, ~ 850 receivers

Optimization strategy: Multiscale, frequency range: 3 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Source/receiver configuration: 100 sources, ~ 850 receivers
Optimization strategy: Multiscale, frequency range: 3 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
**Numerical examples – Marmousi model**

**Source/receiver configuration:** 100 sources, ~ 850 receivers

**Optimization strategy:** Multiscale, frequency range: 3 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Source/receiver configuration: 100 sources, ~ 850 receivers
Optimization strategy: Multiscale, frequency range: 3 Hz to 14 Hz [2 sweeps], Algorithm: L-BFGS (10 iters)
Conclusions and road ahead

Reconstruction algorithm potentially apt to large 3D problems:

- based on “partial” projection of slack variables
- computational properties: can scale to 3D (unlike WRI!), but 2X FWI
- reconstruction quality: more robust to local minima wrt FWI, but inferior results compared to WRI

What’s next:

- time-domain implementation (almost ready)
- TTI acoustic (M. Louboutin)
- implement constraints, checkpointing, …
Time-domain implementation details: Devito/JUDI

Devito:

domain specific language for stencil-based finite-difference C code generation for PDEs w/ explicit time stepping in Python using SymPy

[Luporini et al., 2018; Louboutin et al., 2018]  https://www.devitoproject.org

JUDI:

Julia Devito inversion framework: Julia package based on Devito, high-level abstraction of the linear algebra involved in FWI, WRI, … (data vectors, restriction/injection operators, wave equation solution, forward modeling Jacobian and relative adjoint, …)

[Witte et al., 2019]  https://github.com/slimgroup/JUDI.jl
Open source frequency-domain implementation

Frequency-domain implementation in Julia:

https://github.com/slimgroup/Software.rizzuti2019SEGadf
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)