Compound Regularization of Full-waveform Inversion for Imaging Piecewise Media

H. S. Aghamiry\textsuperscript{1,2}, A. Gholami\textsuperscript{1}, S. Operto\textsuperscript{2},

\textsuperscript{1} Univ. of Tehran, Institute of Geophysics, Tehran, Iran
\textsuperscript{2} Univ. Côte d’Azur, CNRS (IRD), Géoazur, Valbonne, France,

Workshop on Machine-Learning-Assisted Image Formation
10-12 Jul 2019 Nice (France)
What is Full waveform inversion (FWI)?

FWI → Build high resolution earth parameters models (velocity, density, anisotropy, quality factor) with recorded seismograms (Virieux and Operto, 2009).
What is Full waveform inversion (FWI)?

FWI → Build high resolution earth parameters models (velocity, density, anisotropy, quality factor) with recorded seismograms (Virieux and Operto, 2009).

<table>
<thead>
<tr>
<th></th>
<th>Seismic exploration</th>
<th>Medical ultrasound (soft tissues/cortical bone)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave-speed</td>
<td>3000 m/s</td>
<td>1540 / 3500 m/s</td>
</tr>
<tr>
<td>Frequency</td>
<td>2.5-25 Hz</td>
<td>1-5 MHz</td>
</tr>
<tr>
<td>Investigated depth</td>
<td>10 Km</td>
<td>150 mm in soft tissues</td>
</tr>
<tr>
<td>Attenuation</td>
<td>0.5 dB per wavelength</td>
<td>0.1 / 5 dB per wavelength</td>
</tr>
</tbody>
</table>
FWI from mathematical point of view

Unknovns \( \rightarrow \) \( \begin{align*} u &\rightarrow \text{wavefield} \\ m &\rightarrow \text{model parameters} \end{align*} \)

Knowns \( \rightarrow \) \( \begin{align*} d &\rightarrow \text{recorded data} \\ b &\rightarrow \text{source} \end{align*} \)

Global optimization methods

Gradient based methods

FWI \( \rightarrow \) find \( m \) and \( u \)

\( \begin{align*} Pu &= d \\ A(m)u &= b \end{align*} \)

Solve FWI using alternating direction method of multipliers (ADMM) (Boyd et al., 2010)

ADMM based FWI or IR-WRI (Aghamiry et al., 2019e)
2004 BP model: *True model*
Initial model
Classical approach: Estimated model
Classical approach: Estimated model
Regularization

- Due to the **ill-posedness** of FWI, **regularization** is necessary.
- It is used to inject **prior knowledge** and statistical properties to the inversion.
- But, how to choose regularization?

Two popular regularizations in geophysics and image denoising:

1. **Second-order Tikhonov** regularization
   \[
   \|m\|_{Tikh2} = \sum \|\nabla_x^2 m\|_2^2 + \|\nabla_z^2 m\|_2^2.
   \]
   Drive inversion toward **smooth** reconstruction.

2. **Blockiness-promoting Isotropic Total Variation (TV)** regularization
   \[
   \|m\|_{TV} = \sum \sqrt{|\nabla_x m|^2 + |\nabla_z m|^2}.
   \]
   Drive inversion toward piecewise homogeneous (**blocky**) reconstruction.

\(\nabla_i\) and \(\nabla_i^2\): first and second-order difference operators in the \(i\) direction (\(i \in \{x, z\}\)).
When the model become complicated, they can’t recover all the elements.
2D inversion test: What is the shortage of Tikhonov and TV?

True model

Without reg.

Tikhonov

TV
2D inversion test: What is the shortage of Tikhonov and TV?

How we can reconstruct such complicated models?

Either we can use more complicated regularization functions or combine the simple regularizations.
Compound regularization: (Benning and Burger, 2018; Aghamiry et al., 2019b)

Compound regularizers are constructed by combining two or more separate simple regularizers.

- **Convex Combination (CC)**: The solution is forced to satisfy the individual priors simultaneously.

  \[
  \text{Reg}(m) = \alpha_1 \Phi_1(m) + \ldots + \alpha_r \Phi_r(m),
  \]

- **Infimal Convolution (IC)**: The solution is explicitly decomposed into simple components, each of them being regularized by an appropriate prior.

  \[
  \text{Reg}(m) = \min \{ \alpha_1 \Phi_1(m_1) + \ldots + \alpha_r \Phi_r(m_r) \}.
  \]
Compound regularization:

Geometrical illustration of $l_1$, $l_2$, their CC and IC regularizations (Aghamiry et al., 2019b)

Figure 1: (a) the $l_1$-norm, (b) the $l_2$-norm, (c) the CC of $l_1$ and $l_2$-norm, and (d) IC of $l_1$
Compound regularization:

Geometrical illustration of $l_1$, $l_2$, their CC and IC regularizations (Aghamiry et al., 2019b)

What do we want to do?

We are going to apply combination of regularizations in the framework of ADMM-based FWI or IR-WRI (Aghamiry et al., 2019e).

Figure 1: (a) the $l_1$-norm, (b) the $l_2$-norm, (c) the CC of $l_1$ and $l_2$-norm, and (d) IC of $l_1$
Of course at the **global minimum** $u$ and $m$ satisfy the wave-equation as well as the observation equation. FWI with $\text{Reg}(m)$ as a regularization reads

$$\min_{m,u} \text{Reg}(m) \quad \text{subject to} \quad \begin{cases} A(m)u = b \\ Pu = d \end{cases}$$

(4)
• **IR-WRI** → Solving for primals and duals in alternating mode with ADMM.
We test the following regularizers

- **Joint Tikhonov-TV regularizer (JTT):** CC of second-order Tikhonov and first-order TV

  \[
  \text{Reg}(\mathbf{m}) = \alpha_1 \|\mathbf{m}\|_{TV} + \alpha_2 \|\mathbf{m}\|_{Tikh2}.
  \]  
  \[ (7) \]

- **Tikhonov-TV regularizer (TT):** IC of second-order Tikhonov and first-order TV (Gholami and Hosseini, 2013)

  \[
  \text{Reg}(\mathbf{m}) = \min_{m = m_1 + m_2} \left\{ \alpha_1 \|\mathbf{m}_1\|_{TV} + \alpha_2 \|\mathbf{m}_2\|_{Tikh2} \right\}.
  \]  
  \[ (8) \]

- **Total Generalized Variation regularizer (TGV):** IC of first-order and second-order TV (Bredies et al., 2010; Setzer et al., 2011)

  \[
  \text{Reg}(\mathbf{m}) = \min_{m = m_1 + m_2} \left\{ \alpha_1 \|\mathbf{m}_1\|_{TV} + \alpha_2 \|\mathbf{m}_2\|_{TV2} \right\}.
  \]  
  \[ (9) \]
TT Regularized IR-WRI:

Solving the parameter estimation subproblem with TT regularization (Aghamiry et al., 2018, 2019b)

TT regularized and bound constrained parameter estimation subproblem

\[
m^{k+1} = \underset{m \in \mathcal{C}}{\arg \min} \alpha_1 \|m_1\|_{TV} + \alpha_2 \|m_2\|_{Tikh2} + \lambda_1 \|Lm - y\|_2^2,
\]

(10)

where \(m \in \mathcal{C}\) → a bound on the lower and upper bound of the model → \(\{m_l \leq m \leq m_u\}\).

Splitting to solve problem 10

Defining auxiliary variables → \(q\) for bound constraint and \(p\) for TV regularization.

\[
\min \quad \alpha_1 \|p\|_{TV} + \alpha_2 \|m_2\|_{Tikh2} + \lambda_1 \|L[m_1 + m_2] - y\|_2^2 \quad \text{subject to} \quad \begin{cases} 
\ p = \nabla m_1 \\
\ q = m_1 + m_2 \\
\ m_l \leq q \leq m_u
\end{cases}
\]

(11)

where \(\nabla = \begin{bmatrix} \nabla_x \\ \nabla_z \end{bmatrix}\).

Then, we build augmented Lagrangian and use ADMM to solve it.
**ADMM to solve problem 11**

1. For $\mathbf{m} \rightarrow \min_{\mathbf{m}_1, \mathbf{m}_2} \alpha_2 \| \mathbf{m}_2 \|_{\text{Tikh}^2} + \lambda_1 \| \mathbf{L}[\mathbf{m}_1 + \mathbf{m}_2] - \mathbf{y} \|_2^2 + \gamma_0 \| \mathbf{m}_1 + \mathbf{m}_2 - \mathbf{q} - \hat{\mathbf{q}} \|_2^2 + \gamma_1 \| \nabla \mathbf{m}_1 - \mathbf{p} - \hat{\mathbf{p}} \|_2^2$,

   $\mathbf{m}_1$ and $\mathbf{m}_2$ can be updated simultaneously using a variable projection scheme without extra computational burden compare to TV regularization (Aghamiry et al., 2019d).

2. For $\mathbf{q}$ (a projection problem) $\rightarrow \min_{\mathbf{q} \in \mathcal{C}} \| \mathbf{m}_1 + \mathbf{m}_2 - \mathbf{q} - \hat{\mathbf{q}} \|_2^2 = \min (\max (\mathbf{m}_1 + \mathbf{m}_2 - \hat{\mathbf{q}}), \mathbf{m}_l, \mathbf{m}_u)$,

3. For $\mathbf{p}$ (a proximity problem) $\rightarrow \min_{\mathbf{p}} \alpha_1 \| \mathbf{p} \|_{TV} + \gamma_1 \| \nabla \mathbf{m}_1 - \mathbf{p} - \hat{\mathbf{p}} \|_2^2 = \text{prox}_{\gamma_1/\alpha_1} (\nabla \mathbf{m}_1 - \hat{\mathbf{p}})$,

4. For $\hat{\mathbf{p}}$ & $\hat{\mathbf{q}}$ (dual ascent) $\rightarrow \begin{cases} \hat{\mathbf{q}} \leftarrow \hat{\mathbf{q}} + \mathbf{q} - \mathbf{m}_1 - \mathbf{m}_2, \\ \hat{\mathbf{p}} \leftarrow \hat{\mathbf{p}} + \mathbf{p} - \nabla \mathbf{m}_1, \end{cases}$

$\gamma_0$, $\gamma_1$ and $\gamma_2$ are penalty parameters.
Some guidelines for tuning parameters of Regularized IR-WRI

<table>
<thead>
<tr>
<th>Penalty param.</th>
<th>$\alpha_1$ &amp; $\alpha_2$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraints</td>
<td>Compound reg. weight</td>
<td>Obs. Eq.</td>
<td>Wave Eq.</td>
<td>TV weight</td>
<td>Bounds</td>
</tr>
</tbody>
</table>

**Table 1:** $\alpha_1$ & $\alpha_2$: balance different regularization functions. $\lambda_1$, $\lambda_2$, $\gamma_1$, $\gamma_0$: weights of the observation equation, wave equation, auxiliary TV term, bound constraint wrt. regularization term, respectively.

Some guidelines to select penalty parameters (Aghamiry et al., 2019c)

- We found that $\alpha_1 = 0.7$ & $\alpha_2 = 0.3$ was a good pragmatical value.
- We use $\gamma_0 = \gamma_1$.
- We found that $\gamma_1/\alpha_1 = 2\% \max \sqrt{|\nabla_x m_1 - \hat{p}_1|^2 + |\nabla_z m_1 - \hat{p}_2|^2}$ was a good pragmatical value.
- $\gamma_1/\lambda_1$: small percentage of mean absolute value of the diagonal coefficients of $L^T L$.
- $\lambda_1/\lambda_2$: small percentage of of the highest eigenvalue of $A(m)^{-T} P^T P A(m)^{-1}$ (van Leeuwen and Herrmann, 2016; Aghamiry et al., 2019e)
Zero offset VSP test:

IR-WRI with Noiseless data - 100 iterations
Regularized IR-WRI: Application to the BP salt model (left target)

Experimental setup

- Fixed-spread surface acquisition.
- Frequency bandwidth: 3-13 Hz.
- Frequency continuation: Batches of 3 frequencies with a 0.5Hz spacing.
- Three paths over batches.
- Noiseless data.
- Stopping criterion of iteration:

\[ k_{\text{max}} = 15 \quad \text{or} \quad (\|A(m)u - b\|_F \leq 1e - 3 \quad \text{and} \quad \|Pu - d\|_F \leq 1e - 5), \]

where \( k_{\text{max}} \) is the maximum number of iterations.
Regularized IR-WRI: *First frequency batch after 45 iterations*
Source (left) and data (right) residuals at first iteration (top) and at convergence point (bottom)
Regularized IR-WRI: *Final models*
Regularized IR-WRI: Final models
Regularized IR-WRI: Final models
Regularized IR-WRI: Final models
Regularized IR-WRI: Final models

<table>
<thead>
<tr>
<th>Regularizer</th>
<th># iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMP</td>
<td>426</td>
</tr>
<tr>
<td>Tikhonov</td>
<td>448</td>
</tr>
<tr>
<td>TV</td>
<td>399</td>
</tr>
<tr>
<td>JTT</td>
<td>415</td>
</tr>
<tr>
<td>TT</td>
<td>361</td>
</tr>
<tr>
<td>TGV</td>
<td>394</td>
</tr>
</tbody>
</table>

**Table 2:** Number of iterations of IR-WRI for each regularizer.
Conclusions

- We have proposed a versatile recipe to cascade bound constraints and various regularizations in ADMM-based WRI (IR-WRI).
- Nonsmooth regularization is easily implemented with the so-called split Bregman method and proximal algorithms.
- The subsurface is formed by different components of different statistical properties. Need to combine different regularizations.
- These regularizations should be combined by infimal convolution rather than by convex combination.
We thank SEISCOPE consortium and their sponsors for their continuous support.

- This study was partially funded by the SEISCOPE consortium (http://seiscope2.osug.fr), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL.
- This study was granted access to the HPC resources of SIGAMM infrastructure (http://crimson.oca.eu), hosted by Observatoire de la Côte d’Azur and which is supported by the Provence-Alpes Côte d’Azur region, and the HPC resources of CINES/IDRIS/TGCC under the allocation 0596 made by GENCI.


TT regularized IR-WRI: Application of ADMM (or Split Bregman)

- Primal descent

\[
\begin{bmatrix}
m_1^{k+1} \\
m_2^{k+1}
\end{bmatrix} = \arg \min_{m_1, m_2} C(m_1, m_2, p^k, m^k, p^k, \tilde{m}^k),
\]

\[
p^{k+1} = \arg \min_p \alpha \|p\|_1 + \frac{\zeta}{2} \|\nabla m_1^{k+1} - p - \tilde{p}^k\|_2^2,
\]

\[
m^{k+1} = \arg \min_{m \in C} \frac{\eta}{2} \|m_1^{k+1} + m_2^{k+1} - m - \tilde{m}^k\|_2^2,
\]

where

\[
C(m_1, m_2, p^k, m^k, \tilde{p}^k, \tilde{m}^k) = \frac{\gamma}{2} \|L[m_1 + m_2] - y\|_2^2 + (1 - \alpha) \|\nabla^2 m_2\|_2^2
\]

\[
+ \frac{\zeta}{2} \|\nabla m_1 - p^k - \tilde{p}^k\|_2^2 + \frac{\eta}{2} \|m_1 + m_2 - m^k - \tilde{m}^k\|_2^2,
\]

- Dual ascent

\[
\tilde{p}^{k+1} = \tilde{p}^k + p^{k+1} - \nabla m_1^{k+1},
\]

\[
\tilde{m}^{k+1} = \tilde{m}^k + m^{k+1} - (m_1^{k+1} + m_2^{k+1}).
\]
TT regularized IR-WRI: Subproblem \((m_1,m_2)\) - Jointly updating \(m_1\) and \(m_2\) by variable projection

- \((m_1,m_2)\) are solution of the following system

\[
\begin{bmatrix}
G_{11} & G_{12} \\
G_{21} & G_{22}
\end{bmatrix}
\begin{bmatrix}
m_1 \\
m_2
\end{bmatrix}
=
\begin{bmatrix}
h_1 \\
h_2
\end{bmatrix},
\]

with

\[
\begin{aligned}
G_{11} &= \gamma L^T L + \zeta \nabla^T \nabla + \eta I \\
G_{12} &= G_{21} = \gamma L^T L + \eta I \\
G_{22} &= \gamma L^T L + (1 - \alpha)(\nabla^2)^T \nabla^2 + \eta I
\end{aligned}
\]

and

\[
\begin{aligned}
h_1 &= \gamma L^T y + \zeta \nabla^T [p^k + \tilde{p}^k] + \eta [m^k + \tilde{m}^k] \\
h_2 &= \gamma L^T y + \eta [m^k + \tilde{m}^k]
\end{aligned}
\]

where \(I\) is the identity matrix.

- From the first equation of (15), we find that

\[
m_2 = G_{12}^{-1} [h_1 - G_{11} m_1]
\]

and plugging this into the second equation of (15) we get the following

\[
m_1 = (G_{11} - G_{22} G_{12}^{-1} G_{11})^{-1} [h_2 - G_{22} G_{12}^{-1} h_1].
\]

Interestingly, \(L\) is diagonal, implying that \(G_{12}\) is also diagonal. Thus we only need to solve an \(n \times n\) system to estimate \(m_1\), from which \(m_2\) easily follows.
TT regularized IR-WRI: \textit{Subproblem (p) and (m) - Proximity operators}

- \( p = [p_x \ p_z]^T \) estimated with a generalized proximity operator (Combettes and Pesquet, 2011)

\[
p^{k+1} = \text{prox}_{\zeta/\alpha}(z) = \begin{bmatrix} \xi \circ z_x \\ \xi \circ z_z \end{bmatrix},
\]

where

\[
z = \nabla m_1^{k+1} - \hat{p}^k = \begin{bmatrix} z_x \\ z_z \end{bmatrix},
\]

and

\[
\xi = \max \left( 1 - \frac{\zeta}{\alpha \sqrt{z_x^2 + z_z^2}}, 0 \right).
\]

- The subproblem for \( m \) has also a component-wise solution given by

\[
m^{k+1} = \text{proj}_C(m_1^{k+1} + m_2^{k+1} - \tilde{m}^k),
\]

where the projection operator projects its argument onto the desired box \([m_l, m_u]\) according to

\[
\text{proj}_C(\bullet) = \min(\max(\bullet, m_l), m_u).
\]
TT regularized IR-WRI: Final models (Logs)
**Projection subproblem** → \( q = \arg \min_{q \in \mathcal{C}} \| \bullet - q \|^2_2 = \text{proj}_\mathcal{C}(\bullet) = \min(\max(\bullet, m_i), m_u) \)

Approximates the input point with some other point in the desired set \( \mathcal{C} \) which is closest to it in the L2 sense.

**Proximity subproblem** → \( p_1 = \arg \min_{p_1} \sum \sqrt{|p_1|^2 + |p_2|^2} + \gamma_1 \| \nabla_x m_1 - p_1 - \hat{p}_1 \|^2_2 = \text{prox}_{\gamma_1}(\nabla_x m_1 - \hat{p}_1) = \max(1 - \frac{1}{\gamma_1 \sqrt{\| \nabla_x m_1 - \hat{p}_1 \|^2 + \| \nabla_z m_1 - \hat{p}_2 \|^2}}, 0)[\nabla_x m_1 - \hat{p}] \)

Approximates the input point with some other point closest to it in the L2 distance sense under regularization implemented with the penalty term \( \sum \sqrt{|p_1|^2 + |p_2|^2} \).