A denoising formulation of Full-Waveform Inversion

Rongrong Wang and Felix J. Herrmann,
Seismic Laboratory for Imaging and Modelling, University of British Columbia

ABSTRACT

We propose a wave-equation-based subsurface inversion method that in many cases is more robust than the conventional Full-Waveform Inversion. The new formulation is written in a denoising form that allows the synthetic data to match the observed ones up to a small error. Compared to the Full-Waveform Inversion, our method treats the noise arising from the data measuring/recording process and that from the synthetic modelling process separately. Comparing to the Wavefields Reconstruction Inversion, the new formulation mitigates the difficulty of choosing the penalty parameter λ. To solve the proposed optimization problem, we develop an efficient frequency domain algorithm that alternatively updates the model and the data. Numerical experiments confirm strong stability of the proposed method by comparisons between the results of our algorithm with that from both plain FWI and a weighted formulation of the FWI.

INTRODUCTION

All wave-equation based seismic inversion techniques suffer from both additive noise and modelling errors. Albeit smaller, the latter can cause more damage to the inverted model. Depending on specific experiment settings, modelling errors arising from PDE discretization, the use of inaccurate modelling kernels, trace truncations, timing errors, source estimation and location errors all contribute to the noise in quite different ways.

These modelling errors are not taken into account in conventional Full-Waveform Inversion (FWI) formulations, where the PDE constraints are strictly imposed (Tarantola and Valette, 1982; Virieux and Operto, 2009),

$$\min_{m} \sum_{i}^{ns} \|P_{\Omega_i} u_i - d_i\|_2^2$$
$$\text{subject to } A(m)u_i = q_i, \quad i = 1, \ldots, ns.$$

Following the usual notation, we used q_i to denote the ith source, d_i to denote the corresponding data, P_{Ω_i} to denote the restriction operator to receiver locations, and A for the discretized Helmholtz matrix for the model m at a specific frequency. Although the strict PDE constraints increases the numerical efficiency by allowing one to eliminate u_i and derive the reduced form,

$$\min_{m} \sum_{i}^{ns} \|P_{\Omega_i} A(m)^{-1} q_i - d_i\|_2^2$$

it is not very effective in handle modelling errors. Note that modelling errors may be generated from all regions of the model. They become part of the data after propagating to the receiver locations. As noise terms, they are often more coherent to than pure additive noise, therefore the two-norm data misfit may not be an appropriate way to model them. As is indeed observed, FWI is very robust to white noise but less so to other types of noises such as those coming from the source.

A better way of treating the modelling error is to allow a small misfit in the PDE, like the formulation in the Wavefield Reconstruction Inversion (WRI) (van Leeuwen and Herrmann, 2013a, 2015)

$$\min_{m, u_i, i=1,\ldots,ns} \sum_{i}^{n_s} \|P_{\Omega_i} u_i - d_i\|_2^2 + \lambda \|A(m)u_i - q_i\|_2^2.$$

Despite the similarity to FWI, both the data and model misfits are now softly penalized in their ℓ_2 norms with certain weight λ containing prior information of their relative strengths van Leeuwen and Herrmann (2013b). However, setting the parameter λ is a problem, as the modelling error is often unknown. In contrast, the energy of pure data side noise is easier to evaluate by data processing techniques as such noise is usually close to being Gaussian. Note that we consider interfering signals from unknown sources as modelling error here, the data side noise only consists of those introduced by the measuring procedure of the receivers.

Recognizing this relative difficulty of estimating model-side errors and the relative easiness of estimating data-side errors, we hereby propose a denoising formulation, named FWI-DN (denoising), as an alternative to FWI with enhanced stability

$$\min_{m,u} \sum_{i} \|\mathcal{D}_z(A(m)u_i - q_i)\|_2^2 \qquad \text{(FWI-DN)}$$
$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \le \epsilon_i, i = 1, \ldots, n_s$$

Here ϵ_i is the estimated data-side noise in ℓ_2 norm for the ith shot gather, and $\mathcal{D}_z$ is a linear operator that performs depth weighting on the modelling error with weights that are nondecreasing with depth. In this way, we avoid the difficulty to estimate the modelling error, and at the same time acquire the flexibility of incorporating different noise levels for different sources.

We end this section by addressing the related weighted FWI. One may argue that the usual FWI formulation can also be modified to handle the non-uniform noise case where different sources or even traces have different noise levels, through introducing weights to the FWI misfit e.g., Farquharson and Oldenburg (1998). Specifically, we can reformulate the objective as

$$\min_{m} \sum_{i} w_i \|P_{\Omega_i} A(m)^{-1} q_i - d_i\|_2^2 \qquad (1)$$

where a natural choice of the weight could be $w_i = \epsilon_{i_0}^2 / \epsilon_i^2$ for some fixed i_0 and for all $i = 1, \ldots, n_s$. This way, the shot gather corresponding to a larger noise level ϵ_i is relatively lightly penalized, reflecting the correct believe that this source is less reliable. However, putting the misfits for all sources in a mixed objective form causes this formulation fail to provide guarantees for the final solution to lie inside the ϵ_i-ball of the real data. Furthermore, due to the nonlinearity of the objective function, the weighted FWI can get trapped at local minima before entering the region where the misfits become proportionally to the weights. To further support these arguments, a comparison of the performances between the weighted FWI and the proposed method will be presented in the numerical section.

METHODOLOGY

Solving the denoising problem

To solve FWI-DN, we propose to use the alternative minimization method. For numerical efficiency considerations, instead of updating the wavefields u_i ($i = 1, ..., ns$) and the model m alternatively, we suggest to do the alternative update between the data $P_{\Omega_i} u_i$ ($i = 1, ..., ns$) and the model m. To be more specific, let us first write out the subproblems. At the k-th iteration, for fixed m^k , we obtain $P_{\Omega_i} u_i^{k+1}$ by solving a quadratic constraint problem

$$P_{\Omega_i} u_i^{k+1} = \arg\min_{u_i} \|\mathcal{D}_z(A(m^k)u_i - q_i)\|_2^2, \qquad (2)$$
$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \le \epsilon_i.$$

Then fix $P_{\Omega_i} u_i^{k+1}$, and solve for m^{k+1} through (3),

$$(m^{k+1}, \widetilde{u}^{k+1}) = \arg\min_{m, u_1, \ldots, u_{n_s}} \sum_i \|\mathcal{D}_z(A(m)u_i - q_i)\|_2^2,$$
$$\text{subject to } P_{\Omega_i} u_i = P_{\Omega_i} u_i^{k+1}, \qquad (3)$$

where $\widetilde{u}^{k+1} = [\widetilde{u}_1^{k+1}, ..., \widetilde{u}_{n_s}^{k+1}]$ represents a collection of the n_s wavefields that minimizes (3).

Intuitively, each iteration of subproblem (2) can be interpreted as a data denoising procedure with the output $P_\Omega u_i^{k+1}$ being the denoised data. The denoised data is then fitted exactly in solving the subproblem (3). The denoising step will get increasingly more accurate as the model iterates m^k become better.

Notice that subproblem (3) can be written into an unconstrained formulation

$$(m^{k+1}, \widetilde{v}^{k+1}) \qquad (4)$$
$$= \arg\min_{m, v_1, \ldots, v_{n_s}} \sum_{i=1}^{n_s} \|\mathcal{D}_z\{A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T P_{\Omega_i} u_i^{k+1}) - q_i\}\|_2^2,$$

where Ω_i^c is the complementary set of Ω_i and $P_{\Omega_i}^T$ is the transpose of the operator P_{Ω_i}. More specifically, (4) is obtained by setting a new variable v_i through $v_i = P_{\Omega_i^c} u_i$ and substituting the u_i in (3) by

$$u_i = P_{\Omega_i}^T P_{\Omega_i} u_i + P_{\Omega_i^c}^T P_{\Omega_i^c} u_i$$
$$= P_{\Omega_i}^T P_{\Omega_i} u_i^{k+1} + P_{\Omega_i^c}^T v_i.$$

where we have used the constraint in (3). Note that since the variables $\tilde{v}_i^{k+1}$ will not be used in future iterations, the whole purpose of solving subproblem (4) is to update the model m.

The algorithm now reduces to solving (2) and (4) alternatively. As we will see in the next section, solving the subproblem (2) for the wavefields u_i to a relatively high accuracy typically involves several inversions of the augmented system $[A; \lambda P_\Omega]$ for some λ. Subproblem (4) can be solved using variable projection as in the Wavefields Reconstruction Inversion whose cost is proportional to the number of model updates multiplied by the cost of augmented system $[A; \lambda P_\Omega]$ inversion. We found that the most efficient way to solve the whole (FWI-DN) problem is to solve (2) with high accuracy during each iteration followed by performing a few updates of the subproblem (4) for v_i and m, so that the two subproblems have similar computational cost.

Some readers may wonder that instead of (4), why not use a more natural way of updating m, that is to fix u_i^{k+1} and to minimize m via

$$m^{k+1} = \arg\min_m \sum_i \|\mathcal{D}_z(A(m)u_i^{k+1} - q_i)\|_2^2, \qquad (5)$$

The reason we prefer (4) to (5) is that when solving (5) we cannot make much progress in updating m since the complete wavefields remain fixed, whereas this is better in (4) because only the wavefields at the boundary are fixed.

Solving the denoising problem

The subproblem (2) is a least-squares problem with a norm inequality constraint. It is known that for this type of problem, u_i can be evaluated by transforming it to a related Lasso problem, i.e., the penalty formulation of WRI with a specific parameter λ, and be solved by carrying out one inversion of the augmented system $[A; \lambda P_\Omega]$.

Specifically, to solve subproblem (2), we use the Lagrangian dual approach, which solves the dual problem

$$\max_{\lambda \ge 0} G(\lambda) \qquad (6)$$

where G is the Lagrange dual of (2) and λ is the dual variable, i.e.,

$$G(\lambda) = \min_{u_i} \|\mathcal{D}_z(A(m)u_i - q_i)\|_2^2 + \lambda \|P_{\Omega_i} u_i - d_i\|_2^2 - \lambda \epsilon_i.$$

Since G is differentiable with respect to λ, one can solve $G'(\lambda) = 0$ for the minimizer $\hat{\lambda}$ and use the strong duality (More, 1993) to obtain u_i. It is easy to calculate that

$$G'(\lambda) = \|P_{\Omega_i} \bar{u}_i(\lambda) - d_i\|_2^2 - \epsilon_i$$

with $\bar{u}_i$ being the solution to the related Lasso problem

$$\bar{u}_i = \arg\min_{u_i} \|\mathcal{D}_z(A(m^k)u_i - q_i)\|^2 + \lambda \|P_{\Omega_i} u_i - d_i\|_2^2, \quad (7)$$

which has the closed form solution

$$\bar{u}_i = \begin{bmatrix} \mathcal{D}_z(A(m^k)) \\ \sqrt{\lambda} P_{\Omega_i} \end{bmatrix}^\dagger \begin{bmatrix} \mathcal{D}_z(q_i) \\ \sqrt{\lambda} d_i \end{bmatrix}$$

where $\dagger$ is the pseudo inverse. Also, we can obtain the second derivative

$$G''(\lambda) = -2(P_{\Omega_i} \bar{u}_i - d_i)^T P_{\Omega_i} C^{-1} P_{\Omega_i}^T (P_{\Omega_i} \bar{u}_i - d_i),$$

where

$$C = A(m)^T \mathcal{D}_z^T \mathcal{D}_z A(m) + \lambda P_{\Omega_i}^T P_{\Omega_i}.$$

The equation $d'(\lambda) = 0$ can then be solved by the Newton's method. Start with $\lambda = \lambda_0$ and update

$$\lambda^{k+1} = \lambda^k - G'(\lambda)/G''(\lambda).$$

According to the strong duality principle (More, 1993), once the optimal λ^* is found, we can obtain u_i^{k+1} through (7).

Updating the model

To solve (4), we use the variable projection approach as in WRI. For a fixed m, v_i^{k+1} has a unique closed form optimal solution that minimizes the objective function in (4)

$$\bar{v}_i = \mathcal{D}_z(A_{\Omega_i^c})^\dagger(m^k)\mathcal{D}_z(q_i + A_{\Omega_i}(m^k)P_{\Omega_i}u_i^k),$$

where $A_{\Omega_i^c}(m^k)$ denotes the submatrix of $A(m^k)$ formed by columns indexed by Ω_i^c. Using this to project out v_i and rewriting (4) as

$$m^{k+1} = \arg\min_m \|\mathcal{D}_z\{A(m)(P_{\Omega_i^c}^T \bar{v}_i + P_{\Omega_i}^T P_{\Omega_i} u_i^{k+1}) - q_i\}\|_2^2,$$

it is easy to calculate its gradient g and Gauss-Newton Hessian H with respect to m. As mentioned before, for numerical efficiency we only perform a few updates of m We hence update m by the Newton's step

$$m^{k+1} = m^k - H^{-1}(m^k)g(m^k). \tag{8}$$

Now we summary the main steps of the proposed algorithm.

Algorithm 1 Algorithm for solving WRI-DN

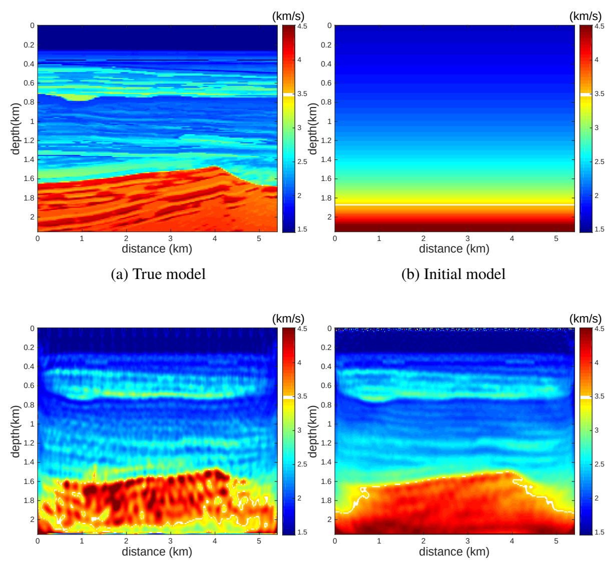
```
procedure INPUT: d_i, Ω_i, A, m^0, T_1, T_2
    for frequency = f_low, ...., f_high do
        for k = 1,... T_1 do
            for each source i ← {1, ..., n_s} do
                update u_i^k by solving (2)
            end for
            for j= 1,..., T_2 do
                updating m using (5).
            end for
        end for
    end for
end procedure
```

NUMERICAL EXPERIMENT

Experiment 1

We test the robustness of the proposed method in the presence of modelling errors. We use a special type of modelling error coming from inaccurate estimates of the source signature. The test model is the 2D BG-compass model. Sources and receivers are placed at 12 m in depth with source spacing 240 m and receiver spacing 48 m. All sources have the same Ricker wavelet signature centred at 10 Hz. The wrongly estimated source signature is a shrinkage of the true one by 20%. We compare the results of the proposed method with that of the conventional FWI. For our method, since only modelling errors exist in this case, we set $\epsilon_i = 0$ and the output of subproblem (2) is therefore simply $P_{\Omega_i} d_i$. As a result, for each frequency, we only need to solve the subproblem (4) without any alternating. We choose a simple linear depth weighting $\mathcal{D}_z = z$, which we observed greatly enhanced the stability of our algorithm compared to using a constant weight. The inversions are performed in frequency domain one frequency slice at a time from 2Hz to 15 Hz. For each frequency, the minimization problems for both methods are solved until convergence.

The inverted results from FWI and our method are shown in Figure 1c and Figure 1d, respectively. Compared to the true model in Figure 1a, both inversions are kinematically correct. However, our result is more accurate in high-frequency reconstruction, therefore has greater smoothness. The final model reconstruction error of our method is about 60% of that from the FWI. We also observe that at the shallow part of the inverted model especially around the receiver locations, our result is very noisy. This is a consequence of the increased depth weighting, which pushes the modelling error up to the shallow part and guarantees the deep part to be stable. Since the deep part is usually the region of interest, this is a benefit of incorporating the weight.



(a) True model (b) Initial model

(c) Inverted model with FWI (d) Inverted model with FWI-DN

Figure 1: A comparison of stability of FWI with FWI-DN under source estimation error: (a) The true model; (b) The initial model; (c) Inversion result with FWI; (d) Inversion result with the proposed method FWI-DN and linear depth weighting.

Experiment 2

We test the robustness our method with respect to strong white Gaussian noise. The SNR of the low-frequency data (2 – 10 Hz) used in this example is 0 dB, and that for high-frequency (10 – 15 Hz) is 25 dB. We use shallow water Marmousi model with 50 m water layer. Sources and receivers are placed at 12m depth with source and receiver spacing of 240 m and 48 m, respectively. All the inversions are performed in the frequency domain one frequency slice at a time from 2 Hz to 15 Hz.

We assume non-uniform noise. The data associated with sources located on the left half of the top model are polluted by white Gaussian noise at level ϵ, and those on the right half are damped by the same type of noise at a different levels 3ϵ. In the FWI-DN, we set the corresponding noise threshold to be 0.8ϵ and 2.4ϵ which are inaccurate and conservative estimates of the true noise levels. We comment that it is always preferable to choose a conservative estimate as we do not need the output of the first subproblem to be completely noise free. This is because the second subproblem (4) itself is also a stable algorithm that can handle the rest of the noise. Besides, if the noise threshold ϵ_i is set too high, then the algorithm will remove a significant portion of the signal components as well, leading to an unsatisfactory update of the model.

For comparison, we run both our algorithm and the weighted FWI. For the latter, we set weights to exactly reflect the prior information that sources on the left are 3 times more accurate than those on the right. Specifically, let N_1 be the set of source indices on the left and N_2 be those on the right. The weighted FWI minimizes

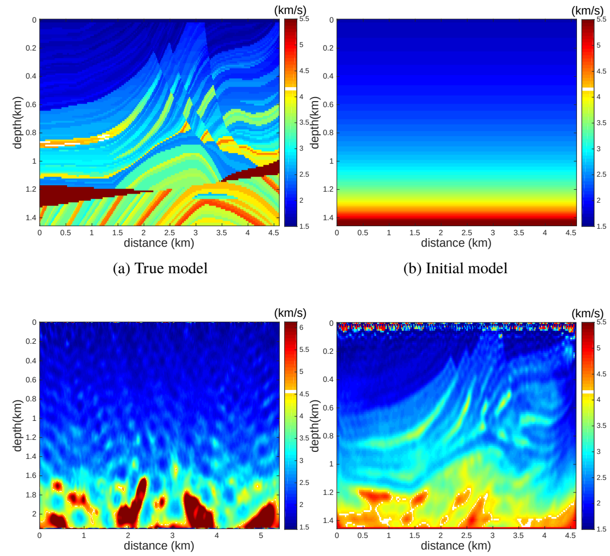
$$\min_m \sum_{i\in N_1} 9\|P_{\Omega_i} A^{-1}(m)q_i - d_i\|_2^2 + \sum_{i\in N_2} \|P_{\Omega_i} A^{-1}(m)q_i - d_i\|_2^2.$$

Figure 2c shows the inverted model using weighted FWI started from the initial guess displayed in Figure 2b. We observed that the noise is too large for the weighted FWI to even keep the kinematic correctness of the inversion. In contrast, our method not only produces a kinematically correct model but is also able to filter out the noise and keep the smoothness in the reconstruction. Admittedly, the deep part of the model is not well reconstructed, due to the fact that the weak signal components coming from the deep part are filtered out along with the noise. This seems to be an inevitable consequence in such a high noise scenario. Moreover, as in the previous example, there are large inversion errors in the shallow water layer region, this is due to the depth weighting which allows a larger modelling error for the shallower part than the deeper part. Fortunately, this error does not propagate down to affect the reconstruction of those regions that we are most interest in.

Finally, it is worth mentioning that one can modify the proposed method into a pure noise attenuation method by deactivating its model updating step (3). We use the initial guess m_0 as a tuning parameter of the denoiser, and solves

$$\bar{u} = \arg\min_{u_i, i=1,\ldots,n_s} \sum_i \|\mathcal{D}_z(A(m_0)u_i - q_i)\|_2^2$$
$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \le \epsilon_i, i = 1, \ldots, n_s.$$

The output of the algorithm $P_{\Omega_i}\bar{u}_i$ is then a smooth approximation to d_i. The smoothness is a result of minimizing the PDE misfit, as wavefields that obey the PDE has a certain level of regularity. If the noise level ϵ_i is too small, then the output is close to the input. On the other hand, if ϵ_i is too larger, say larger than $\|d_i - d_i^0\|_2$ with d_i^0 being the data of the initial model, then $\bar{u}$ is simply the wavefiled of m_0, so we lose all the



(a) True model
(b) Initial model
(c) Inverted model with FWI
(d) Inverted model with FWI-DN

Figure 2: A comparison of robustness of FWI with FWI-DN under Gaussian noise with SNR=0 dB: (a) The true model; (b) The initial model; (c) Inversion result with weighted FWI; (d) Inversion result with the proposed method FWI-DN and linear depth weighting.

information in the data about $\delta m = m_{true} - m_0$. When ϵ is neither too large nor too small, the method smooths the data towards the direction of the initial guess. Hence the better the initial guess is, the better the denoiser performs. In addition, when the computational speed is a concern, one can sacrifice some accuracy and use a very small m_0, say only the top water layer.

DISCUSSION AND CONCLUSIONS

We proposed a denoising formulation for FWI named FWI-DN, which separately treats the data-side and model-side noises as opposed to combing them into one objective as in the conventional FWI. The WRI formulation is in between the proposed method and the FWI, and is equivalent to our problem if an oracle is given for the parameter λ that balances the data and the PDE misfits. In a sense, our formulation is an extension of the WRI formulation. We proposed an efficient algorithm for FWI-DN which performs alternating updates on the model and the data. The part of the algorithm that solves the data updating subproblem can be used as a stand-alone denoising method when good initial guesses of the model is available.

ACKNOWLEDGEMENT

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.

REFERENCES

Farquharson, C. G., and D. W. Oldenburg, 1998, Non-linear inversion using general measures of data misfit and model structure: Geophysical Journal International, **134**, 213–227.

More, J. J., 1993, Generalizations of the trust region problem.: Optimization methods and Software, 189–209.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232.

van Leeuwen, T., and F. J. Herrmann, 2013a, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

———, 2013b, A penalty method for pde-constrained optimization.

———, 2015, A penalty method for PDE-constrained optimization in inverse problems: Inverse Problems, **32**, 015007. ((Inverse Problems)).

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.