Preconditioning FWI with the sparse additive inverse laplacian correlation function

P. Wellington*, † R. Brossier*, O. Hamitou†, S. Garambois*, J. Virieux*
*ISTerre Univ. Grenoble Alpes, France; †Laboratoire Jean Kuntzmann (LJK), Univ. Grenoble Alpes, France
†Speaker

SUMMARY

Full waveform inversion (FWI) is a non-linear, ill-posed, local data fitting technique, that can benefit significantly from preconditioning and/or regularization. We present an efficient 2D preconditioning operator, constructed from combining two orthogonal 1D inverse wavenumber correlation operators. Our sparse inverse operator is used to perform local dip-variant wavenumber filtering by calculating the solution of a linear system. To highlight the utility of this approach, we apply it to two FWI case studies: the synthetic Marmousi model and a marine streamer dataset from the North Western Continental shelf of Australia.

INTRODUCTION

Full waveform inversion (FWI) is an increasingly popular data fitting technique that allows one to move from an initial low wavenumber representation of the subsurface to a more complete and accurate broadband-wavenumber image (see Virieux and Operto, 2009, for a review). FWI considers either the complete and accurate broadband-wavenumber image (see Virieux, 1984) or a windowed subset, as it looks to minimize the difference between the observed field data \(d_{\text{obs}}\) and the modelled data \(d_{\text{mod}}\). To maximize efficiency of the data fitting goal, a local linearized optimization is considered from a sufficiently accurate starting model (Tarantola, 1987).

As the FWI problem, is a non-linear, ill-posed problem, our model parameter reconstruction can suffer from undesired imprints from factors such as the true nature of wave propagation, the limited frequency content in the inversion, the presence of noise and/or illumination issues. A number of strategies can be used to attempt to mitigate against these problems. Common examples include the addition of a penalty term based on roughening operators (Press et al., 1986) to emphasize smoothness (Tikhonov and Arsenin, 1977) or edges (Guittion et al., 2012). Alternatively, approaches using model re-parametrization methods (Clapp et al., 2004; Fomel and Claerbout, 2003) can be applied to FWI, (Guittion et al., 2012; Lewis et al., 2014) where local dip filters based on anisotropic diffusion kernels (Hale, 2007) allows one to take benefit of prior knowledge of the interpreted dip of expected reflection events.

In this abstract, we detail an alternative strategy where the FWI gradient is preconditioned using a correlation operator. The correlation function acts as a local wavenumber filter and is constructed by a novel and new approach involving a multi-dimensional extension of the 1D inverse laplacian correlation function (Tarantola, 1987).

METHODOLOGY

FWI is a data-fitting process that looks to update the model parameters, \(m\), involved in wave propagation, so that the modelled data \(d_{\text{mod}} = d(m)\) matches the observed field data \(d_{\text{obs}}\). The full waveform inversion typically minimizes the misfit based on an \(\ell_2\) objective function

\[ C(m) = \frac{1}{2} || d_{\text{obs}} - d_{\text{mod}} ||^2. \]

Starting from an initial model \(m_0\), we update the current model \(m_n\) at the iteration \(n\) with a perturbation model \(\Delta m_n\). This defines the updated model

\[ m_{n+1} = m_n + \alpha \Delta m_n, \]

where \(\alpha\) represents the step length taken along \(\Delta m_n\). The perturbation should have contributions from the first (gradient) and second derivative (hessian) of the objective function. In our approach, we calculate the gradient using the adjoint state method (Plessix, 2006) and converge to a solution using the quasi-newton L-BFGS scheme (Brossier et al., 2009; Nocedal, 1980). Further details of the local optimization utilized can be found in Métivier and Brossier (2016)

Gradient preconditioning

In order to limit the influence of undesired features such as noise, we would like to precondition the inversion using a spatial parameter correlation description. This can be thought as a local 2D low-wavenumber filter that can be used to prevent non-geological features entering into the model \(m_{n+1}\). The correlation operator defines how perturbations at a point \((x, z)\) are correlated to perturbations at another point \((x', z')\). An example of such a function is the 2D laplacian correlation function

\[ \text{Corr}_{2D}(x, z; x', z') = \frac{1}{2 \pi L_x L_z} e^{- \frac{1}{2} \sqrt{ (x-x')^2 + (z-z')^2 } } \]

where \(L_x\) and \(L_z\) are correlation lengths in the horizontal and vertical directions. Such a correlation operator can be combined with a rotation matrix to allow anisotropic filtering with the long correlation length \(L_{\text{cor}}\) parallel to the local structural dip, \(\theta\) and a short correlation length \(L_{\text{cor}}\) orthogonal to dip (Figure 1).

We would like to replace our standard FWI gradient

\[ G_{\text{cal}} = \frac{\partial C(m_n)}{\partial m} \]

to a preconditioned descent direction

\[ G'_{\text{cal}} = \text{Corr}_{2D} G_{\text{cal}}. \]

Unfortunately, the operator \(\text{Corr}_{2D}\) is quite large (the same dimension as the mono-parameter Hessian) and banded as it

© 2016 SEG
SEG International Exposition and 86th Annual Meeting
FWI application of sparse laplacian correlation function

Figure 1: The correlation lengths can either be aligned with the cartesian coordinates ($L_x$ and $L_z$) or alternatively be rotated by a local angle $\theta$, ($L_{x\text{rot}}$ and $L_{z\text{rot}}$).

describes how a given point in the gradient $g_i$ correlates to all other surrounding points. The computation of $G'$ is represented by an area integral for every element in the gradient vector, which is computationally expensive to perform.

Efficient sparse inverse Laplacian covariance

The 1D Laplacian correlation function

$$\text{Corr}_{1D}(z,z') = \frac{1}{2\pi} e^{-\frac{|z-z'|}{\epsilon}}$$ (6)

is quite unique as there is an analytical expression to the inverse operator given by (Tarantola, 1987)

$$\text{Corr}_{1D}^{-1}(z,z') = Lh \left( \frac{1}{L} \delta(z-z') - L\delta^2(z-z') \right).$$ (7)

where $L$ is our 1D correlation length, $h$ is the grid cell size and the terms $\delta(z-z')$ and $\delta^2(z-z')$ are a delta function and the second derivative of the delta function respectively. We attempted to extend expression 7 to 2D by adding two orthogonal 1D inverse correlation functions through

$$\text{Corr}_{2D\text{tot}}^{-1}(x,x',z,z') = L_xh \left( \frac{1}{L_x} \delta(x-x') - L_x\delta^2(x-x') \right) + L_zh \left( \frac{1}{L_z} \delta(z-z') - L_z\delta^2(z-z') \right).$$ (8)

We refer to that operator as the additive inverse laplacian (AIL) correlation function. One can discretize $\delta(x-x')$, $\delta(z-z')$, $\delta^2(x-x')$ and $\delta^2(z-z')$ using finite differences and, doing so, we obtain a 5-points stencil for applying the correlation function to unrotated correlation lengths ($L_x$ and $L_z$) while we require a 9-points stencil if we would like our 2 orthogonal correlation lengths ($L_{x\text{rot}}$ and $L_{z\text{rot}}$) to be rotated by a dip, $\theta$. The sparse nature of $\text{Corr}_{2D\text{tot}}^{-1}$, as well as the diagonal dominant structure of the operator, means that one can solve the linear system

$$A \cdot \text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D} \approx u$$ (9)

to efficiently calculate $\text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D} \cdot u$. Solving this linear system once does not appear to approximate $\text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D} \cdot u$. However, if we solve two successive linear systems (Algorithm 1), we obtain a quite fair approximation of $\text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D} \cdot u$ at a low computational cost. To further highlight the behavior of our sparse linear system, we investigate two arbitrary vectors. These vectors, $u$ are of length $101 \times 101$. One example shows a dirac spike ($u = \delta$) and the second is a white noise vector ($u = \epsilon_w$) (Figure 2). Anisotropic correlation lengths are utilized, where $L_{x\text{rot}} = 4L_{z\text{rot}}$ and $\theta = 45^\circ$. The similarity of the result of solving the linear system twice to the analytical laplacian correlation function

$$x_{2D} = \text{Corr}_{2D\text{tot}}^{-1}(\text{Corr}_{2D\text{tot}}^{-1} \cdot u) \approx \text{Corr}_{2D\text{tot}}^{-1} \cdot u.$$ (10)

is a interesting result, that I will explore in subsequent work. In this abstract to minimize over-filtering orthogonal to geological dip in our FWI examples we will use the preconditioned gradient

$$G' = \text{Corr}_{2D\text{tot}}^{-1} \cdot G,$$ (11)

which only requires the solution of one linear system.

**Algorithm 1** Approximate $\text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D}$ with $\text{Corr}_{2D\text{tot}}^{-1} \cdot u$

1. $A \cdot \text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D} \approx u$
2. $A \cdot \text{Corr}_{2D\text{tot}}^{-1} \cdot u \approx \text{Corr}_{2D\text{tot}}^{-1} \cdot x_{2D}$

---

© 2016 SEG
SEG International Exposition and 86th Annual Meeting
FWI application of sparse laplacian correlation function

Figure 3: LEFT: The Marmousi synthetic $v_p$ model, CENTRE: Initial FWI model RIGHT: Approximate dip field of the true model.

FWI EXAMPLES

We will provide two examples illustrating the usefulness of the additive inverse laplacian correlation function to FWI. The first is performed on a subset of the Marmousi synthetic model, while the second is performed on a marine streamer dataset taken from the North Western Shelf of Australia.

Marmousi synthetic example

We perform an application of our preconditioned FWI to a subset of the Marmousi model (Figure 3). The model is composed of 4 rotated fault blocks, with an average dip of approximately 45°. The true model is discretized using a 10m grid and is composed of 130 cells in the vertical and 390 in the horizontal direction. An initial model is built by applying a 500 m isotropic gaussian slowness smoother to the true model. After performing this smoothing, the inclined layers are largely absent from the initial model. We apply frequency-domain full waveform inversion in an attempt to recover this high wavenumber content. A fixed spread acquisition is employed with receivers every 10 m and sources every 50 m. Our FWI application utilizes a frequency sweeping strategy (Bunks et al., 1995), where we invert over 4 successive frequency bands (4 Hz to 10 Hz, 4 Hz to 15 Hz, 4 Hz to 20 Hz and 4 Hz to 25 Hz) using a frequency increment on 1 Hz. White noise is added to the observed seismic data.

The additive inverse laplacian correlation function will mitigate the impact of noise mapped into the current model, $m_n$. We use an anisotropic ($L_{x_{rot}} \neq L_{z_{rot}}$) description of the correlation length where $L_{x_{rot}} \approx 75 m$ and $L_{z_{rot}} \approx 10 m$. This is combined with a spatially variant description of the local dip field (Figure 3). We see that, when no preconditioning is applied to our inversion, our FWI gradient contains the true geological detail in addition to some random noise (Figure 4). By applying the preconditioning based on our additive inverse laplacian, we are able to precondition the gradient by filtering gradients approximately along geodetic dip. This decreases the imprint of the noise that is evident in the final inversion results without removing the geological information.

Real Data example: Offshore Australia

The second example we considered is a 2D real dataset that transects the North-Western Australian Continental shelf. The transects the North-Western Australian Continental shelf. The

Figure 4: The initial gradient from the first iteration of the first frequency band of the inversion TR: without preconditioning BL: with preconditioning. The final inversion results after FWI TR: without preconditioning BR: with preconditioning.

line is acquired with the Broadseis (CGG) seismic acquisition strategy and has a significant variation in water depth from less than 100m to greater than 1000m. This feature combined with other issues in this zone causes significant distortion to the seismic image especially at the exploration target level (approximately 3km depth). An initial velocity model with a grid cell size of 25m (31.25km long and 8km deep) is constructed using sparse RMS stacking velocities built as part of a pre-stack time migration imaging work-flow (see Wellington et al., 2015, for more information on the initial model). A time domain full waveform inversion is applied using a multiscale work-flow (Bunks et al., 1995), where we invert over 6 second-order butterworth frequency bands. The lowest frequency introduced into the inversion is $\approx 1 Hz$ while the highest is $\approx 12.5 Hz$. At the begining of the inversion for each frequency band, source estimation is performed using the entire shot record and, then, velocity and density are both inverted simultaneously using a L-BFGS algorithm. The inverted velocity result from each band serves as the input to the subsequent frequency band while the density model at each frequency band is initialized from a constant density model. Using density in this way allows an effective means for inverting for the density contrast at the sea floor which we expected to be quite rugose. The inversion results (Figure 5) has additional high wavenumber information in the near surface when compared to the initial model. This results in a significant reduction in distortion of the seismic image (Figure 6). One of the issues noticed in the unpreconditioned FWI results is that there is some vertically oriented and high-wavenumber noise in the final velocity reconstruction. These details are perhaps easier to see in Figure 7. The information is not consistent with what should be

© 2016 SEG
SEG International Exposition and 86th Annual Meeting
FWI application of sparse laplacian correlation function

Figure 5: The NWA line focuses on the water column and near surface. From the initial model (TOP) we perform unpreconditioned FWI (CENTER) and we can see high-wavenumber information being introduced to the inversion especially very close to water-bottom. Our preconditioned FWI (BOTTOM) is able to remove this information.

expected in the region where the geology dip trends in the near surface in a horizontal/sub-horizontal fashion. We employ a preconditioner using $L_x = 100\, \text{m}$ and $L_z = 12.5\, \text{m}$ (no dip included) to attenuate this information while looking to preserve the vertical wavenumber information.

CONCLUSION

We have introduced a new and efficient preconditioning operator for the FWI gradient, that is calculated by leveraging the 1D inverse laplacian correlation function. With limited prior information and limited parameterization, we are able to stabilize FWI from some of the ill-posed issues that will map into the model-space. Further real data examples, extensions to 3D and other uses of the additive inverse laplacian operators will be discussed in future studies.

ACKNOWLEDGEMENTS

We thank CGG for access to the Australian real dataset as well as providing processing support and free access to a Hampson Russel license. This study was partially funded by the SEISCOPE consortium (http://seiscope2.osug.fr), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE.

Figure 6: CIGs calculated using the (TOP) initial velocity model and (BOTTOM) the velocity model after FWI. The green square highlights a gas field at 3km depth. Notice the improvement of gather alignment on the FWI CIGs.

Figure 7: The difference $\Delta v_p$ between our initial model and FWI results for the result with no preconditioning (TOP) and with preconditioning (BOTTOM).
REFERENCES


