

DEVELOPMENT IN INVERSE HESSIAN ESTIMATION

Least-squares migration (LSM) is a well-established technique for improving the quality of seismic imaging through inversion (??). By formulating seismic imaging as a linear estimation problem, least-squares migration utilizes the power of iterative inversion for recovering the reflectivity of subsurface structures. In recent years, least-squares imaging, in particular LSRTM (least-square reverse-time migration) has found many successful applications (???).

The least-squares formulation involves the inverse Hessian operator. The exact computation of such inverse is prohibitively expensive. In the conventional approach, the inverse Hessian is approximated by iterative methods, such as conjugate gradients (???). This raises the cost of LSRTM to the cost of migration and modeling multiplied by the number of iterations. Only a small number of iterations can be affordable in practice.

To reduce this cost or, alternatively, to reduce the number of iterations by accelerating the iterative convergence, a number of methods have been proposed in the literature for approximating the inverse Hessian using relatively inexpensive computations. ? estimated a simple scaling operator to approximate the inverse Hessian with a diagonal matrix. ? and ? extended this approach to matching filters, approximating the inverse Hessian with a nonstationary convolution operator. This approach is also known as *migration deconvolution* (??). ? approximated inverse Hessian as a deblurring filter for regularization and precondition scheme. ? adopt the deep learning technique and use generative adversarial networks in a conditional setting (CycleGANs) to approximate inverse Hessian.

An asymptotic theory (??) shows that the inverse Hessian can be represented as a combination of weights in the time domain and the frequency domain. The result extends to the case of LSRTM (??). Motivated by this theory, we propose to approximate the inverse Hessian by reformulate by simultaneously estimating a chain of weights in the space and frequency domains from initial images. We design this chain to be symmetric to preserve the symmetric property of the Hessian. Once the weight matrices are estimated, they can be efficiently approximate the inverse Hessian. Alternatively, the approximate inverse Hessian can be incorporated in the form of a preconditioner to accelerate the convergence of iterative LSRTM.

The paper is organized as follows: We first present the theory, problem formulation, and an estimation algorithm. Next, we test the accuracy of the proposed approach using zero-offset synthetic data from the Marmousi model (?). Subsequently, we use prestack Marmousi data to evaluate the effectiveness of the proposed approximation as a migration-deconvolution filter and as a preconditioner for LSRTM.

THEORY

Problem formulation

Given surface seismic data \mathbf{d} , the first migrated image \mathbf{m}_1 is obtained by applying seismic migration as the adjoint of forward linear modeling operator \mathbf{L} :

$$\mathbf{m}_1 = \mathbf{L}^T \mathbf{d} . \quad (1)$$

Next, we can model and migrate again, generating the second image

$$\mathbf{m}_2 = \mathbf{L}^T \mathbf{L} \mathbf{m}_1 \quad (2)$$

If we can find an effective approximation for the Hessian operator from equation (??)

$$\mathbf{H} \approx \mathbf{L}^T \mathbf{L} \quad (3)$$

and its inverse, then applying this inverse to the initial image \mathbf{m}_1 will provide an effective approximation for the desired least-squares inverse

$$\mathbf{m}_3 = \mathbf{H}^{-1} \mathbf{m}_1 \approx \left(\mathbf{L}^T \mathbf{L} \right)^{-1} \mathbf{L}^T \mathbf{d} . \quad (4)$$

Chain operator

We propose to approximate the connection between \mathbf{m}_1 and \mathbf{m}_2 in equation (??) using a chain of weights alternating between the original domain and the Fourier domain, as follows:

$$\mathbf{m}_2 \approx \mathbf{W} \mathbf{F}^{-1} \mathbf{W}_f \mathbf{F} \mathbf{W} \mathbf{m}_1 , \quad (5)$$

where \mathbf{W} and \mathbf{W}_f are diagonal matrices, and \mathbf{F} represents the Fourier transform. Our goal is to estimate \mathbf{W} and \mathbf{W}_f from the given pairs of \mathbf{m}_1 and \mathbf{m}_2 ,

In the digital filtering terminology, the operation of $\mathbf{F}^{-1} \mathbf{W}_f \mathbf{F}$ is equivalent to convolution. Therefore, the proposed chain represents a combination of scaling and convolution in a symmetric setting. The symmetry is important because the Hessian operator $(\mathbf{L}^T \mathbf{L})$, which we are approximating using this representation, is symmetric by definition.

Once we obtain \mathbf{W} and \mathbf{W}_f , inverting the chain is straightforward. The compensated migrated image \mathbf{m}_3 from equation (??) becomes

$$\mathbf{m}_3 = \mathbf{W}^{-1} \mathbf{F}^{-1} \mathbf{W}_f^{-1} \mathbf{F} \mathbf{W}^{-1} \mathbf{m}_1 . \quad (6)$$

Solving for \mathbf{W} and \mathbf{W}_f

To solve the nonlinear representation of chain of operators in equation (??), we introduce intermediate vectors \mathbf{x}_1 and \mathbf{x}_2 and treat them as additional unknown variables,

splitting the initial condition into three linear equations, as follows:

$$\mathbf{x}_1 \approx \mathbf{W}\mathbf{m}_1\mathbf{x}_2 \approx \mathbf{F}^{-1}\mathbf{W}_f\mathbf{F}\mathbf{x}_1\mathbf{m}_2 \approx \mathbf{W}\mathbf{x}_2 \quad (7)$$

In the vector notation,

$$\mathbf{r} = \mathbf{W} - \mathbf{I}000\mathbf{F}^{-1}\mathbf{W}_f\mathbf{F} - \mathbf{I}000\mathbf{W} - \mathbf{I}\mathbf{m}_1\mathbf{x}_1\mathbf{x}_2\mathbf{m}_2 \approx \mathbf{0} , \quad (8)$$

and the solution is found by minimizing the residual vector \mathbf{r} while updating \mathbf{w} , \mathbf{w}_f , \mathbf{x}_1 , and \mathbf{x}_2 . Linearizing equation (??) with respect to perturbations in unknown variables, we arrive at

$$-\mathbf{I}0\text{diag}[\mathbf{m}_1]0\mathbf{F}^{-1}\mathbf{W}_f\mathbf{F} - \mathbf{I}0\mathbf{F}^{-1}\text{diag}[\mathbf{F}\mathbf{x}_1]0\mathbf{W}\text{diag}[\mathbf{x}_2]0\Delta\mathbf{x}_1\Delta\mathbf{x}_2\Delta\mathbf{W}\Delta\mathbf{W}_f \approx -\mathbf{r} \quad (9)$$

We solve (??) using the conjugate-gradient method with shaping regularization that enforces the smoothness of estimated variables (?), then update \mathbf{W} , \mathbf{W}_f , and recalculate residual \mathbf{r} in (??) iteratively until convergence. This corresponds to a regularized Gauss-Newton approach for solving the original nonlinear system.

To initialize the initial time/space weight \mathbf{W} , we use a smooth division of \mathbf{m}_2 by \mathbf{m}_1 (?) and then take a square root. We set the initial frequency weight \mathbf{W}_f to 1.

Preconditioning with chain weights

To improve the results further, we propose to apply the weights \mathbf{W} and \mathbf{W}_f to form a preconditioner to speed up the convergence of iterative least-square migration. The original problem is to solve for \mathbf{m} where $\mathbf{d} \approx \mathbf{L}\mathbf{m}$. We apply the change of variables from \mathbf{m} to \mathbf{y} according to $\mathbf{m} = \mathbf{P}\mathbf{y}$ where \mathbf{P} is the preconditioning matrix. Then the problem becomes

$$\mathbf{d} \approx \mathbf{L}\mathbf{P}\mathbf{y} \quad (10)$$

We then solve this equation for \mathbf{y} with the least-square objective function

$$J(\mathbf{y}) = \frac{1}{2} \|(\mathbf{L}\mathbf{P})\mathbf{y} - \mathbf{d}\|_2^2 , \quad (11)$$

which corresponds to the analytical solution

$$\mathbf{y} = (\mathbf{P}^T\mathbf{L}^T\mathbf{L}\mathbf{P})^{-1}\mathbf{P}^T\mathbf{L}^T\mathbf{d} . \quad (12)$$

After iteratively inverting for \mathbf{y} , we recover the original variable $\mathbf{m} = \mathbf{P}\mathbf{y}$.

According to the approximation of Hessian operator $\mathbf{L}^T\mathbf{L} \approx \mathbf{W}\mathbf{F}^{-1}\mathbf{W}_f\mathbf{F}\mathbf{W}$

$$\mathbf{L}^T\mathbf{L} \approx \mathbf{W}\mathbf{F}^{-1}\mathbf{W}_f\mathbf{F}\mathbf{W} = (\mathbf{P}\mathbf{P}^T)^{-1} , \quad (13)$$

the preconditioner \mathbf{P} can be constructed from chain weights as

$$\mathbf{P} = \mathbf{W}^{-1}\mathbf{F}^{-1}\mathbf{W}_f^{-1/2}\mathbf{F} . \quad (14)$$

This makes the inverted operator in equation (??) close to unitary,

$$(\mathbf{LP})^{\mathbf{T}}\mathbf{LP} \approx \mathbf{I} , \tag{15}$$

which accelerates the convergence when the inversion is carried out by an iterative method.