Accelerating Least Squares Imaging Using Deep Learning Techniques

Janaki Vamaraju
John A. and Katherine G. Jackson School of Geosciences,
The University of Texas at Austin

Jeremy Vila
Shell International Exploration and Production Inc.

Mauricio Araya-Polo [*]
Shell Global Solutions US Inc.

Debanjan Datta
Shell International Exploration and Production Inc.

Mrinal K. Sen
John A. and Katherine G. Jackson School of Geosciences,
The University of Texas at Austin

Abstract

Wave equation techniques have been an integral part of geophysical imaging workflows to investigate the Earth's subsurface. Least-squares reverse time migration (LSRTM) is a linearized inversion problem that iteratively minimizes a misfit functional as a function of the model perturbation. The success of the inversion largely depends on our ability to handle large systems of equations given the massive computation costs. The size of the system almost exponentially increases with the demand for higher resolution images in complicated subsurface media. We propose an unsupervised deep learning approach that leverages the existing physics-based models and machine learning optimizers to achieve more accurate and cheaper solutions. We compare different optimizers and demonstrate their efficacy in mitigating imaging artifacts. Further, minimizing the Huber loss with mini-batch gradients and Adam optimizer is not only less memory-intensive but is also more robust. Our empirical results on synthetic, densely sampled datasets suggest faster convergence to an accurate LSRTM result than a traditional approach.

1 Introduction

Imaging techniques are used in geophysics to produce images of the Earth's subsurface at diverse length scales. These techniques are recently being adopted for monitoring subsurface geological formations that are used for carbon capture and storage (Figure 1). CO_2 sequestration is one of the solutions to the increasing greenhouse gas emissions that cause global warming and climate change. In this context, wave equation based migration algorithms [1, 2] have been an integral part of

[*] Mauricio Araya-Polo is now at Total. All contributions to this work were made while at Shell Global Solutions US Inc.

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

imaging workflows to invert for subsurface properties at greater depths. Among them, least-squares reverse time migration (LSRTM) [3, 4] is the most popular migration method due to its ability to image complex subsurface areas with large computational resources such as graphic processing units (GPUs). LSRTM can be seen as a linear inverse problem based on the acoustic wave equation. The goal is to invert for an earth model that represents rock properties to fit the recorded surface data. The imaging condition or the gradient calculation in LSRTM is to take the zero-lag of the cross-correlation between the reverse-time-extrapolated receiver wavefield and the forward-time-extrapolated source wavefield [5]. When the data is subjected to aliasing, truncation or noise, the adjoint operator can degrade the resolution of the final migrated image. LSRTM obtains an approximate image of the model perturbation by iteratively minimizing the cost function. The LSRTM, however, is an expensive replacement and its ability to migrate surface data depends on the accuracy of the velocity model and adequate preconditioning [6, 7], and needs additional regularization terms for successful damping of artifacts [8, 9]. At each iteration of LSRTM, Born forward modeling and adjoint operators are applied, which makes the computational cost extremely high.

Under the paradigm of theory guided data science, many researchers in the geophysics community are looking at ways of combining physics and machine learning [10–17]. These works, however were not extended to least-squares migrations. In this paper, we implement LSRTM using a deep learning approach and adopt strategies from data science to reduce computational costs and accelerate convergence.The feasibility of LSRTM is examined in combination with mini-batch gradients and deep learning optimizers such as the Hopfield neural networks (HNN), adaptive moment estimation (Adam) and Limited memory BFGS (L-BFGS). Mini-batch gradients help in reduction of cross-talk [18–20] and the deep learning optimizers can help mitigate acquisition footprints that are caused by the lack of shot data. This not only achieves faster convergence through iterations but also generates geologically consistent models. The computation cost is further reduced by using a subset of total shot data for each iteration. Implementing LSRTM in a deep learning framework (Pytorch or Tensorflow) enables us to experiment with machine learning loss functions and regularizations. The automatic differentiation capability of the software can be used to calculate the gradient of the cost function. We further minimize the Huber loss function to improve the efficiency of LSRTM. We apply the techniques to a 2D synthetic model and show improvement over conventional LSRTM baselines. The proposed methodology achieves higher spatial resolution according to quantitative evaluation metrics.

2 Methodology

The goal of least-squares migration is to invert for the earth's surface reflectivity model (model perturbation indicating rock properties) $\mathbf{m}$ to fit the recorded data, $\mathbf{d_0}$:

$$C(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{d_0} - \mathbf{Gm}\right\|^2, (1)$$

where C is the cost/loss function to be minimized and $\mathbf{G}$ is the linearized Born modeling (scattering) operator that requires a background velocity model. This background velocity model is known a-priori from other velocity analysis methods. If $\mathbf{G^T G}$ is invertible, the least-squares solution for equation (1) can be written as:

$$\mathbf{m} = (\mathbf{G^T G})^{-1}\mathbf{G^T d_0}, (2)$$

where $\mathbf{G^T}$ is the migration operator and $\mathbf{G^T G}$ is the Hessian matrix $\mathbf{H}$. The key to LSRTM is to obtain the inverse of $\mathbf{H}$; however the computational cost and storage of $\mathbf{H}$ are not feasible for realistic problems. Alternatively, different approximations, such as gradient based iterative approaches [21, 3, 22] are pursued. In this paper, we use PyTorch [23] for the above implementation. The Born modeling based wave equations can be implemented as a recurrent neural network (RNN) which is similar to the workflow proposed by [17]. The same operations are applied in each cell of an RNN, but the data that the operations act upon changes. Each cell applies the finite-difference convolution operation to propagate forward one time step, by taking the state from the previous cell (the displacement wavefield at adjacent time steps and auxiliary wavefield) and the source amplitude as inputs, and producing the updated state vectors and the current wavefield as outputs. Although, automatic differentiation can be used to calculate the gradient of the cost function, we use the traditional adjoint state method to save GPU memory costs. Easy chaining of operations in PyTorch, enables users to create customized loss functions or apply data normalizing operations. We generate observed data with the full finite difference wave propagation operator. LSRTM intrinsically is an under-determined inverse problem. There are always a limited number of receivers that can cover the subsurface and the source function is a band limiting signal. This can lead to incomplete data in terms of spatial coverage and frequency content. Here, we use the Huber loss function [24, 25] to improve the resolution and robustness of LSRTM:

$$C(\mathbf{m}) = \frac{1}{N}\sum_{i}^{N} C_i(\mathbf{m}) \text{ where } C_i(\mathbf{m}) = \begin{cases} \frac{1}{2}\left\|\mathbf{d_0} - \mathbf{Gm}\right\|^2 & \text{if } |\mathbf{d_0} - \mathbf{Gm}| \leq \varepsilon \\ \varepsilon\,|\mathbf{d_0} - \mathbf{Gm}| - \varepsilon/2 & \text{otherwise} \end{cases}, (3)$$

where i is the corresponding shot record for i-th source and N is the total number of shots. The parameter ε defines a threshold based on the distance between target and prediction. For the example in this paper, we use a default value of 1 for ε. Conventionally, the LSRTM cost function is evaluated at each iteration using all the shot data that is available. The gradient calculation can be very expensive when the number of shots are large (especially in 3D surveys). To reduce the computation costs and to reap the benefits of stochastic optimization, mini-batch gradient methods take a subset of entire shots to construct the objective function and update the model:

$$C(\mathbf{m}) \approx C_k^B(\mathbf{m}) = \frac{1}{|B|}\sum_{i\in B} C_i(\mathbf{m}), (4)$$

where B is a subset of total shots, and $|B|$ is its size. This performs frequent updates with reasonable variance and is faster to converge. The fluctuation of the objective function is not severe. We first divide all the shot data into sequential mini-batches and then randomly shuffle them. Although mini-batches offer a more scalable solution, they can introduce migration artifacts if the data is not sampled effectively. Therefore, we need more efficient optimization algorithms to produce migrated images with higher resolution. First-order gradient based optimization algorithms can be very slow to converge to high accuracy solutions, specially in noisy large-scale datasets. We combine mini-batch gradients with various optimizers such as the Hopfield neural networks (HNN), adaptive moment estimation (Adam) and Limited memory BFGS (L-BFGS). These techniques not only reduces the number of forward solves but also enhances the accuracy of inverted results when compared to traditional LSRTM algorithms (using gradient descent (GD)). We provide qualitative and quantitative comparisons of the above presented methods against conventional LSRTM baseline migrated images. Note that, no preconditioning was used for either conventional LSRTM or for the proposed mini-batch approach.

3 Results and Discussion

The synthetic example is based on a $2D$ slice from the SEG/EAGE $3D$ salt model [26]. This model is 4 km in depth and 12.5 km wide. We set 160 shots and 160 receivers per shot both deployed on the surface. The peak frequency of the source Ricker wavelet is 6 Hz. Figures 2(a), 2(b) and 2(c) show the true velocity model, the background smooth velocity model and the true reflectivity respectively. Figure 3(a) shows a conventional LSRTM image using GD after 20 epochs. The reflectivity is still far from the true reflectivity and the bottom of the salt body is not very clear. Figure 3(b) shows the migrated image from HNN which is slightly improved with a mean structural similarity index (MSSIM) of 0.79. Figures 3(c) and 3(d) show the migrated images from L-BFGS and Adam optimizers respectively. When combined with mini-batch training, not only the true amplitudes are better recovered but also the salt bottom is now clearly resolved. The reflectors at the bottom of the image are enhanced and there is no loss of continuity. The number of shots in each mini-batch and the step-size for model updates are the two important hyperparameters. We randomly reserve twenty-five percent of all the shots that evenly cover the entire survey. This development dataset is used to evaluate the cost function for 25 combinations of batch-sizes and learning rates (Grid-search optimization). The batch size and learning rate that correspond to lowest cost function value is chosen for each optimizer. For 40 shots and Adam optimizer, the best learning rate and batch-size is $1e^{-7}$ and 4 respectively. The migrated images have an MSSIM index greater than 0.8 and the migration

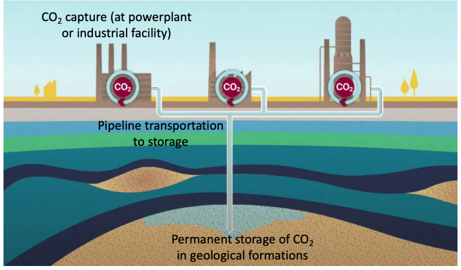
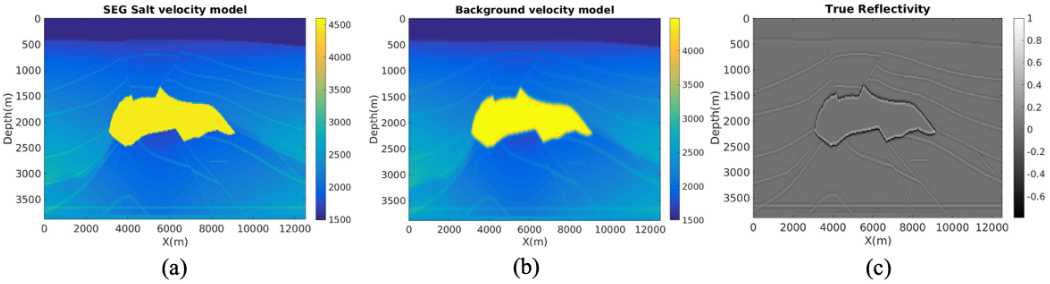

CO_2 capture (at powerplant or industrial facility)
CO_2
CO_2
CO_2
Pipeline transportation to storage
Permanent storage of CO_2 in geological formations

Figure 1: Geophysical imaging of geological formations for enhanced carbon capture and storage



(a)

(b)

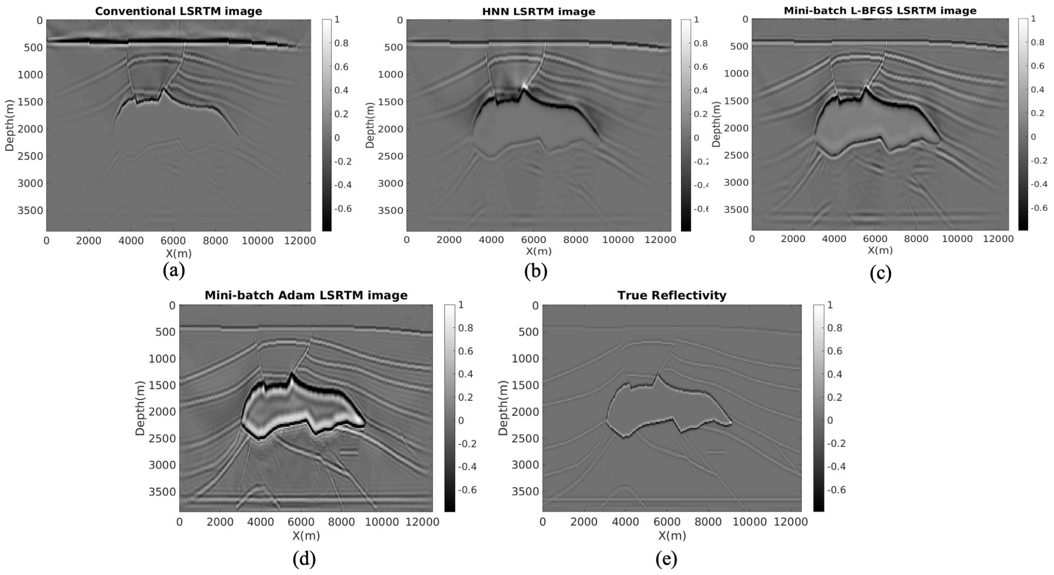
(c)

Figure 2: SEG/EAGE salt model (a) True velocity (b) Smooth migration velocity model (c) True reflectivity

artifacts are removed. Figure 4(a) shows the error convergence or the cost function calculated for each optimizer. Adam and AdaBound converge within 6 epochs and the cost function becomes flat beyond 10 epochs. However, the image from Adam (MSSIM=0.8255) has a higher MSSIM than AdaBound (MSSIM = 0.8157). The performance of L-BFGS is comparable but converges in slightly more number of epochs than Adam. Figures 4(b) and 4(c) plot the other validation metrics for a quantitative evaluation. Using the Adam optimizer, at the end of 20 epochs, MSSIM index reaches a value of 0.8255 from 0.77. Compared to Adam, conventional LSRTM shows almost no improvement from the initial image. Additional preconditioning and filtering is usually needed to greatly improve the image. The final R^2 score of the Adam migrated image is raised to 0.34 from an initial 0.06. The mean squared error (MSE) value is reduced to 0.825 and the peak signal-to-noise ratio (PSNR) is improved to 47.78 over 20 epochs. Note that the PSNR value is measured in log scale. The performance trend is overall positive.

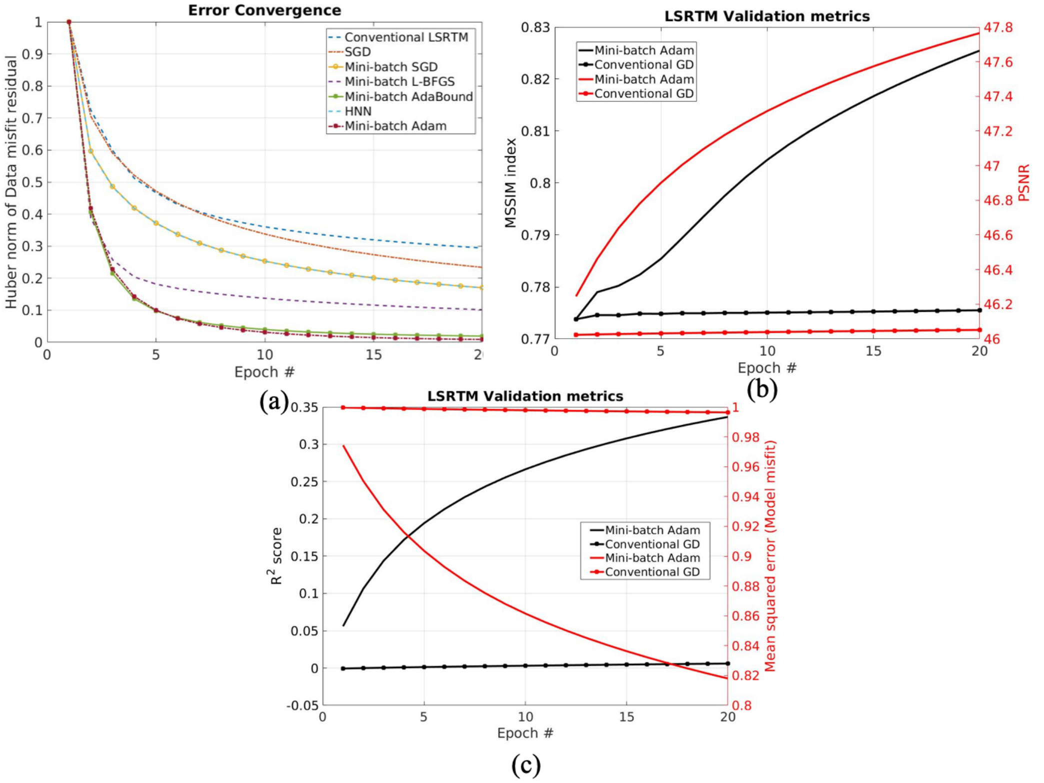
4 Conclusions

We present a deep learning approach for LSRTM by adopting strategies from data-science to accelerate convergence. In a time-domain formulation, mini-batch gradients can reduce the computation cost by using a subset of total shots for each epoch. The Adam optimizer, combined with the Huber loss and mini-batch gradients resulted in significantly faster convergence than when using conventional gradient descent optimizer with the cost and gradient calculated using the entire dataset. We apply the techniques to the SEG/EAGE 3D salt model and show improvements over conventional LSRTM baseline. The described methods mitigate artifacts that arise from limited aperture, low subsurface illumination, and cross-correlation noise. Effects that irregular sampling has on the proposed mini-batch approach remains future work.



(a) (b) (c) (d) (e)

Figure 3: SEG/EAGE salt model migrated images (a) Conventional LSRTM image (MSSIM=0.7755) (b) HNN LSRTM image (MSSIM=0.79) (c) Mini-batch L-BFGS LSRTM image (MSSIM=0.8026) (d) Mini-batch Adam LSRTM image (MSSIM=0.8255) (e) True reflectivity



(a) (b) (c)

Figure 4: Comparison of (a) Error convergence (b) MSSIM and PSNR (c) R^2 and MSE metrics for conventional LSRTM and mini-batch Adam LSRTM images

References

[1] Edip Baysal, Dan D Kosloff, and John WC Sherwood. Reverse time migration. *Geophysics*, 48(11):1514–1524, 1983.

[2] George A McMechan. Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, 31(3):413–420, 1983.

[3] Tamas Nemeth, Chengjun Wu, and Gerard T Schuster. Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221, 1999.

[4] Henning Kühl and Mauricio D Sacchi. Least-squares wave-equation migration for avp/ava inversion. *Geophysics*, 68(1):262–273, 2003.

[5] Jon F Claerbout. Toward a unified theory of reflector mapping. *Geophysics*, 36(3):467–481, 1971.

[6] James E Rickett. Illumination-based normalization for wave-equation depth migration. *Geophysics*, 68(4):1371–1379, 2003.

[7] Antoine Guitton. Amplitude and kinematic corrections of migrated images for nonunitary imaging operators. *Geophysics*, 69(4):1017–1024, 2004.

[8] Zhiguang Xue, Yangkang Chen, Sergey Fomel, and Junzhe Sun. Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization. *Geophysics*, 81(1):S11–S20, 2015.

[9] Di Wu, Gang Yao, Jingjie Cao, and Yanghua Wang. Least-squares rtm with l1 norm regularisation. *Journal of Geophysics and Engineering*, 13(5):666–673, 2016.

[10] Carlos Calderón-Macías, Mrinal K Sen, and Paul L Stoffa. Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation. *Geophysics*, 62(3):992–1002, 1997.

[11] Janaki Vamaraju and Mrinal K Sen. Unsupervised physics based neural networks for seismic migration. *Interpretation*, 7(3):1–51, 2019.

[12] Mauricio Araya-Polo, Taylor Dahlke, Charlie Frogner, Chiyuan Zhang, Tomaso Poggio, and Detlef Hohl. Automated fault detection without seismic processing. *The Leading Edge*, 36(3):208–214, 2017.

[13] Thomas Mejer Hansen and Knud Skou Cordua. Efficient monte carlo sampling of inverse problems using a neural network-based forward—applied to gpr crosshole traveltime inversion. *Geophysical Journal International*, 211(3):1524–1533, 2017.

[14] Mauricio Araya-Polo, Joseph Jennings, Amir Adler, and Taylor Dahlke. Deep-learning tomography. *The Leading Edge*, 37(1):58–66, 2018.

[15] Amir Adler, Mauricio Araya-Polo, and Tomaso Poggio. Deep recurrent architectures for seismic tomography. In *81st EAGE Conference and Exhibition 2019*, 2019.

[16] Reetam Biswas, Mrinal K Sen, Vishal Das, and Tapan Mukerji. Pre-stack and post-stack inversion using a physics-guided convolutional neural network. *Interpretation*, 7(3):1–76, 2019.

[17] Alan Richardson. Seismic full-waveform inversion using deep learning tools and techniques. *arXiv preprint arXiv:1801.07232*, 2018.

[18] Michael P Friedlander and Mark Schmidt. Hybrid deterministic-stochastic methods for data fitting. *SIAM Journal on Scientific Computing*, 34(3):A1380–A1405, 2012.

[19] Tristan van Leeuwen and Felix J Herrmann. Fast waveform inversion without source-encoding. *Geophysical Prospecting*, 61:10–19, 2013.

[20] Hui Yang, Junxiong Jia, Bangyu Wu, and Jinghuai Gao. Mini-batch optimized full waveform inversion with geological constrained gradient filtering. *Journal of Applied Geophysics*, 152:9–16, 2018.

[21] Gerard T Schuster. Least-squares cross-well migration. In *SEG Technical Program Expanded Abstracts 1993*, pages 110–113. Society of Exploration Geophysicists, 1993.

[22] Yaxun Tang. Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian. *Geophysics*, 74(6):WCA95–WCA107, 2009.

[23] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[24] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[25] Antoine Guitton and William W Symes. Robust inversion of seismic data using the huber norm. *Geophysics*, 68(4):1310–1319, 2003.

[26] Fred Aminzadeh, N Burkhard, J Long, Tim Kunz, and P Duclos. Three dimensional seg/eaeg models—an update. *The Leading Edge*, 15(2):131–134, 1996.