Score-based Seismic Inverse Problems

Sriram Ravula[1] **Dimitri Voytan**[1] **Elad Liebman**[2] **Ram Tuvi**[2] **Yash Gandhi**[2]
Hamza Ghani[2] **Alexandre Ardel**[2] **Mrinal Sen**[1] **Alexandros G. Dimakis**[1]
[1]The University of Texas at Austin [2]SparkCognition Research
{sriram.ravula, dvoytan, mrinal}@utexas.edu, dimakis@austin.utexas.edu
{eliebman, rtuvi, ygandhi, hghani, aardel}@sparkcognition.com

Abstract

We present a new family of score-based models designed specifically for seismic migration. We define a sequence of corruptions obtained by migration artifacts created by reverse time migration (RTM) as the number of measurements changes. Our network is conditioned on the number of source locations and refines the reconstructed image over an annealed sequence of steps. Experiments on synthetic seismic data show that we can reconstruct geological details using a very small number of sources. Our method produces significantly higher-quality images compared to posterior sampling using standard score-based generative models and supervised seismic migration baselines.

1 Introduction

Linear inverse problems consist of reconstructing an unknown signal or image $\mathbf{x}^* \in \mathbb{R}^n$ from linear observations, $\mathbf{y} \in \mathbb{R}^m$ where the observations are created by a linear transformation plus noise: $\mathbf{y} = \mathcal{A}(\mathbf{x}^*) + \epsilon$. The forward operator $\mathcal{A}$ is a discretization of the forward physical process that produces the measurements from the unknown system. In this paper we are interested in *seismic imaging* which is modeled by a linear inverse problem where $\mathbf{x}^*$ corresponds to subsurface images of the earth. Seismic imaging is useful for carbon sequestration, understanding the geologic history and the evolution of the Earth, and finding natural resources.

The least squares solution to the linear inverse problem is $\mathbf{x}^* = (\mathcal{A}^\dagger\mathcal{A})^{-1}\mathcal{A}^\dagger\mathbf{y}$, where $\mathcal{A}^\dagger$ is the adjoint of $\mathcal{A}$ and, in the context of seismic imaging, $(\mathcal{A}^\dagger\mathcal{A})^{-1}$ is the inverse Hessian. The Hessian (and particularly its inverse) cannot be realized because each measurement is a collection of time-series that involves typically thousands of samples multiplied by thousands of sensors, requiring terabytes of memory even for a small number of shots. A standard method used in seismic imaging is Reverse Time Migration (RTM) (see e.g. [20]), which circumvents inverting the Hessian with an adjoint approximation. State of the art software like Devito [10, 11] performs RTM by discretizing the partial differential equations of $\mathcal{A}^\dagger$ and computes $\mathbf{y}$ from $\mathbf{x}^*$ through a finite difference solver.

On the other hand, unsupervised deep learning methods for inverse problems have been yielding significant benefits over classical and sparsity-based methods [1, 2, 3, 12, 16]. Unfortunately, it is very challenging to directly apply these methods to seismic imaging since they typically require access to the forward operator.

Our Contribution. We present a new family of score-based models which can be trained from a sequence of baseline reconstructed images with varying numbers of measurements. The baseline reconstructions are created by Devito using RTM and aggregated to create the sequence of images with varying numbers of measurements as we subsequently explain. Our network is conditioned on the number of source locations and refines the baseline RTM image over an annealed sequence of steps. Our experiments show that our annealed process can reconstruct geological details using

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

dramatically fewer measurements (as much as five times) compared to standard RTM as well as posterior sampling with Langevin Dynamics using NCSNv2 [17].

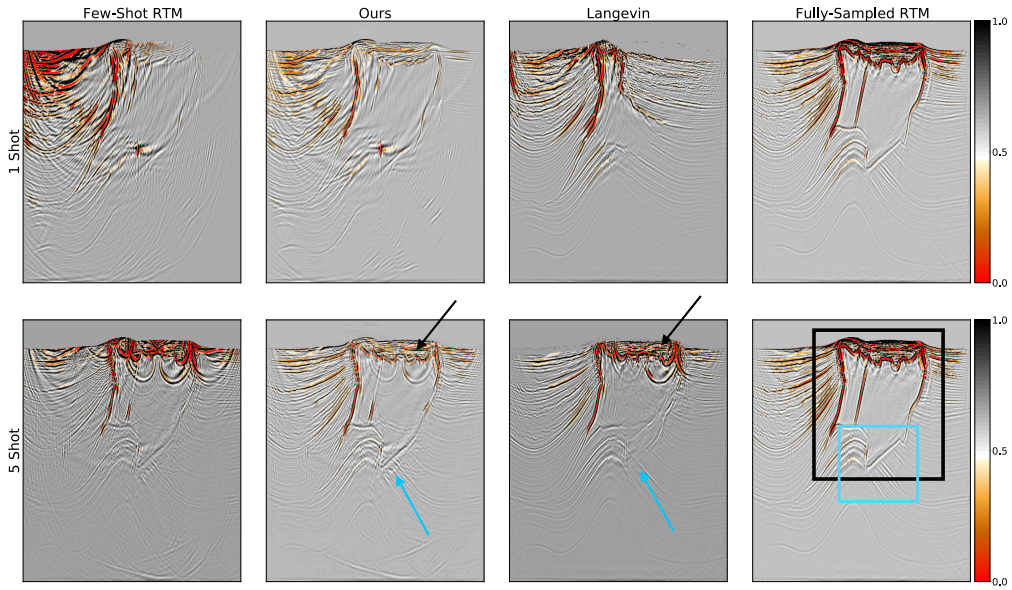

Figure 1: **Visual denoising results.** We denoise migration artifacts from RTM images made with 1 and 5 shots using our method and posterior sampling with NCSNv2 [17]. The fully-sampled RTM images in the right-most column are made with 243 shots, which is the maximum value for all slices in our synthetic dataset. The main feature of interest in this example is the prominent salt structure indicated by the black box in the fully-sampled image in the second row. Remarkably, our method recovers the lower boundary of the salt, highlighted by the light-blue box and light-blue arrows with just 5 measurements, while posterior sampling with Langevin Dynamics creates a poor reconstruction. Furthermore, our method better recovers the top of salt and shallow sediments, indicated by the black arrows, than Langevin sampling.

Background. Marine seismic acquisition involves making many measurements of subsurface echos with instruments placed near the ocean surface. Each measurement is initiated by a towed airgun which generates a pressure pulse. We refer to each measurement as a *source* or *shot*. The instruments record the pulse after it propagates through the Earth, generates reflections from contrasts in geologic layers, and returns to the Earth's surface.

For a set of given measurements $\mathbf{y} \in \mathbb{R}^m$, $m = \ell\tau$ where ℓ is the dimension of measurements for a single source, τ is the total number of sources, and $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_\tau]^T$ is a vector denoting the measurements $\mathbf{y}_i \in \mathbb{R}^\ell$ from individual sources. For each source i we can create a *single-shot* image $\mathbf{x}_i$ using RTM that partially describes the full seismic scene. The true seismic image $\mathbf{x}^*$ we seek to recover can be described simply as the mean of all available single-shot images: $\mathbf{x}^* = \frac{1}{\tau}\sum_{i=1}^{\tau} \mathbf{x}_i$. We can also characterize a partial, or k-*shot* image $\mathbf{x}^k$ made with k out of the available τ shots, where $\mathbf{x}^k = \frac{1}{k}\sum_{j=1}^{k} \mathbf{x}_j$. For each value of k we have $\binom{\tau}{k}$ possible k-shot images, each formed by taking different combinations of k single-shot images.

We can consider the k-shot images to be drawn from a sequence of unknown conditional image distributions $p_k(\mathbf{x}^k|\mathbf{x})$, $k = 1, 2, \ldots, \tau$ where $\mathbf{x} \coloneqq \mathbf{x}^\tau$. Empirically, the more shots included in a partial seismic image, the closer the partial image is to the true image we wish to recover. This is reflected in the sequence of partial distributions, so that as $k \to \tau$ the sequence converges to the true distribution of fully-sampled seismic images $p_\tau(\mathbf{x}^\tau|\mathbf{x}) \coloneqq p(\mathbf{x})$.

Seismic Score Network. Inspired by the success of score-based generative models for solving inverse problems, we propose to train a network with a supervised variant of denoising score matching [18] to recover fully-sampled seismic images. Instead of adding artificial Gaussian noise to create intermediate perturbed data distributions, we estimate the natural sequence of perturbed distributions that results from creating RTM images with varying numbers of measurements. During inference, we

use a variant of annealed gradient ascent in which our trained network traverses from small to large values of k to smoothly remove artifacts from RTM and recover important features.

We train a neural network $\mathbf{f}_\theta(\mathbf{x}^k, k)$ to minimize the following loss:

$$\mathcal{L}(\theta; \tau) = \frac{1}{\tau - 1} \sum_{k=1}^{\tau-1} \lambda(k) \ell(\theta; k), \tag{1}$$

where the loss for each value of k is given by:

$$\ell(\theta; k) = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \mathbb{E}_{\mathbf{x}^k \sim p_k(\mathbf{x}^k|\mathbf{x})} \left[\left|\left| \mathbf{f}_\theta(\mathbf{x}^k, k) + \frac{\mathbf{x}^k - \mathbf{x}}{\sigma^2(\mathbf{x}^k, \mathbf{x})} \right|\right|_2^2 \right]. \tag{2}$$

Here, $\sigma(\mathbf{x}^k, \mathbf{x}) \coloneqq \sqrt{||\mathbf{x}^k - \mathbf{x}||_2^2 / n}$, i.e. the root mean squared error (rmse) between the k-shot image and the fully-sampled image. $\lambda(k) = \sigma^2(\mathbf{x}^k, \mathbf{x})$ are hyperparameters that scale each loss term so their magnitudes are balanced. We also calculate the running empirical mean of the $\sigma(\mathbf{x}^k, \mathbf{x})$ values encountered during training for each value of k and store them as parameters $\{\sigma_k\}_{k=1}^{\tau-1}$. We incorporate the value of k in the input of the network by conditioning the output as $\mathbf{f}_\theta(\mathbf{x}^k, k) \coloneqq \mathbf{f}_\theta(\mathbf{x}^k)/\sigma(\mathbf{x}^k, \mathbf{x})$ for an unconditional network $\mathbf{f}_\theta(\mathbf{x}^k)$. During inference, when we do not have access to the ground truth sample $\mathbf{x}$ we use the mean rmse instead and condition as $\mathbf{f}_\theta(\mathbf{x}^k, k) \coloneqq \mathbf{f}_\theta(\mathbf{x}^k)/\sigma_k$.

Although Eq. (2) is similar to denoising score matching [18], it does not estimate the true score function $\nabla \log p_k(\mathbf{x}^k)$, $p_k(\mathbf{x}^k) \coloneqq \int p_k(\mathbf{x}^k|\mathbf{x}) p_{data}(\mathbf{x})$ since the noise process that generates migration artifacts in few-shot seismic images does not have a known, differentiable density function. Regardless, the loss in Eq. (2) provides a useful signal for the network to learn the best direction to go from a k-shot seismic image with migration artifacts to a fully-sampled image.

During inference, we are given seismic measurements $\mathbf{y}_i = \mathcal{A}(\mathbf{x}^*; i) + \epsilon$, for shot indices $i = 1, 2, \ldots, k_0$, where $k_0 < \tau$. In other words, we are given a subset of the maximum number of available measurements. We form a few-shot image $\mathbf{x}^{k_0}$ by performing RTM on each measurement $\mathbf{y}_i$ and then averaging the resulting images. Letting $\mathbf{x}_0 = \mathbf{x}^{k_0}$ and $k_t|_{t=0} = k_0$, we perform the update:

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t + \alpha_{k_t} \mathbf{f}_\theta(\mathbf{x}_t, k_t). \tag{3}$$

Starting from the initialization, we run T steps of Eq (3) while conditioning the network on k_0 shots. Using the output of that process as the next initialization, we again run T steps while conditioning on $k_0 + 1$ shots. We repeat this process until we iterate T times for $\tau - 1$ shots and output the final image. For each successive noise level, we anneal the step size α_{k_t} at the rate $\epsilon \cdot (\sigma_k^2 / \sigma_{\tau-1}^2)$, where σ_k is the empirical mean of the rmse for k-shot images that we calculated during training.

By traversing the number of shots k, our network iteratively moves the solution closer to the intermediate image distributions $p_{k_0}(\mathbf{x}^{k_0}), p_{k_1}(\mathbf{x}^{k_1}), \ldots, p_{\tau-1}(\mathbf{x}^{\tau-1})$ in sequence. The annealed step size has the effect to dampen updates to the solution as it gets closer to $p(\mathbf{x})$. While the final distribution our algorithm traverses is $p_{\tau-1}(\mathbf{x}^{\tau-1})$ and not $p(\mathbf{x})$, images from each distribution are perceptually indistinguishable for sufficiently large τ.

2 Experimental Results

We show that our proposed method is effective for removing migration artifacts from few-shot RTM images. We build our dataset from SEAM, which is a 3D synthetic environment that mimics a real marine seismic setting [4]. We take 2D slices of the environment and simulate $\tau = 243$ measurements for each slice, then calculate single-shot RTM images for each measurement. The resulting dataset contains 8047 total slices, which we split into a training set with 7847 slices and a test set with 200 slices. The migrated images are single-channel and have a resolution of 751×625.

For all experiments with our method, we use a model trained with the loss in Eq. (1). We perform inference with $T = 2$ and $\epsilon = 2 \times 10^{-8}$. As a baseline, we compare to posterior sampling with Langevin Dynamics using NCSNv2 [17]. Since it is infeasible to differentiate through the forward operator $\mathcal{A}$, we instead model the gradient of the log likelihood as $\nabla_{\mathbf{x}_t} \frac{1}{n} ||\mathbf{x}^k - \mathbf{x}_t||_2^2$ where $\mathbf{x}^k$ is the initial few-shot image and $\mathbf{x}_t$ is our estimate at iteration t. We also scale the log likelihood gradient by a term ζ_t which we anneal as $t \to \infty$, as in previous works [5, 6]. For the architecture of both

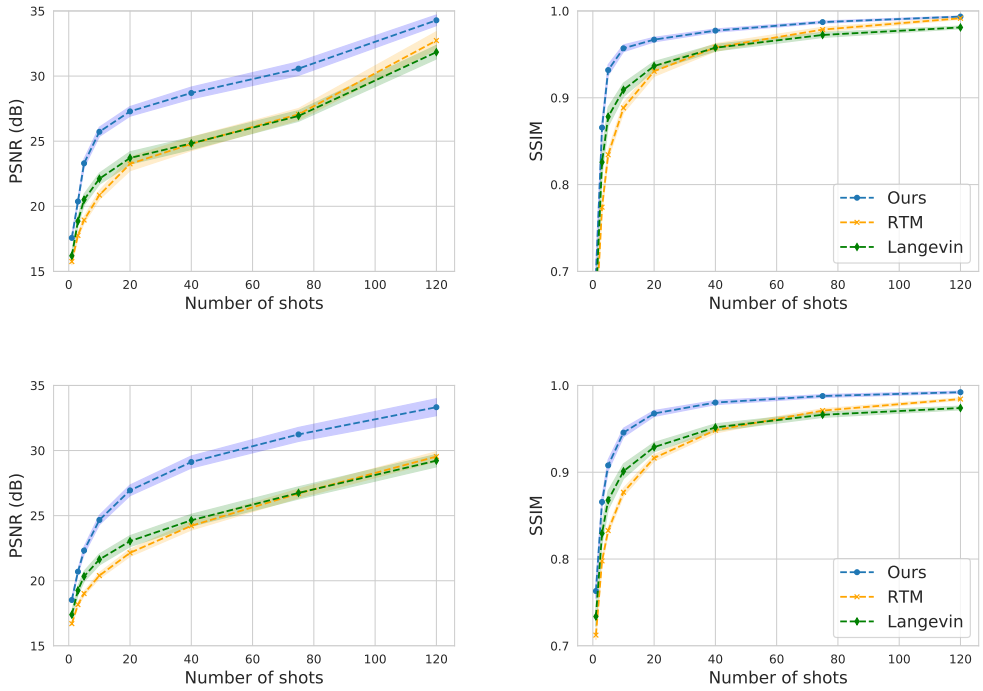

Figure 2: **Quantitative denoising results. Top row: equal source spacing. Bottom row: random source spacing.** We present the mean PSNR and SSIM for denoising migration artifacts from RTM images using our method and posterior sampling with NCSNv2 [17]. Shaded regions indicate 95% confidence intervals. Our method consistently outperforms Langevin Dynamics and improves PSNR and SSIM from the initial image across all shot counts.

our model and the NCSNv2 baseline, we use a variant of the RefineNet [9, 8] presented in [17]. We train both models using the Adam optimizer [7] with a learning rate of 2×10^{-5} and epsilon of 10^{-3}. Both models are trained for 500 epochs with a batch size of 32.

Qualitative Results. In this experiment, we choose a slice from the test set and create 2 RTM images with 1 and 5 shots, respectively. We denoise each image using our method and Langevin Dynamics and present the results in Figure 1. We are interested in recovering the boundaries of the large salt structure indicated in the fully-sampled image. The k-shot images show significant migration swings, which particularly obscure the top, right, and bottom boundaries of the structure.

Our method is able to recover the general outline of the salt structure from just a single shot. Langevin Dynamics removes the significant migration swings in the left half of the single shot image, but reconstructs a non-existent structure and completely misses all but the left boundary of the salt. At 5 shots, our method accurately finds the bottom of the structure whereas Langevin Dynamics attenuates this feature. In general, Langevin Dynamics removes fine migration swings in the shallow sediment to the left and right of the structure better than our method. However, we find that our method outperforms Langevin Dynamics for distinguishing important features in the subsurface from migration swings.

Quantitative Results. In this setting, we create RTM images with $k = 1, 3, 5, 10, 20, 40, 75$, and 120 shots for slices in the test dataset and reconstruct them using our method and posterior sampling with Langevin Dynamics. We perform this process with both randomly-spaced shots and equally-spaced shots. We report the mean PSNR and SSIM [19] values, as well as 95% confidence intervals for the reconstructions and for the initial k-shot images w.r.t. the fully-sampled images in Figure 2.

For both random and equal shot spacing, our method matches the performance of posterior sampling in PSNR and SSIM with up to $5\times$ fewer measurements. Our method exhibits higher variance than posterior sampling in PSNR for high values of k, but lower variance in SSIM across most

measurement ranges. In addition, our method consistently improves on the initial k-shot RTM images, while the reconstruction ability of posterior sampling saturates between $k = 40$ and $k = 75$ and actually produces reconstructions that reduce image quality. Overall, our method outperforms posterior sampling with Langevin Dynamics across all measurement ranges.

Broader Impact

We propose a method for removing migration artifacts from seismic images. Seismic imaging is a crucial tool for geologists and geophysicists to understand regional tectonics and stress fields, relative timings of depositional events, and the evolution of deep-water sedimentary systems [14, 13]. This information is useful for understanding the geological history of a region and the potential for natural disasters such as earthquakes and tsunamis [15], for carbon sequestration, and for finding natural resources. By reconstructing seismic images that capture important geological details with up to $5\times$ fewer measurements than other methods, our method can significantly reduce the cost of measurement acquisition for scientific surveys. For applications in natural resource extraction, the improvement in reconstructed image quality over other methods can lead to a reduction in unnecessary exploration, and therefore a reduction in negative environmental impacts from activity such as drilling.

Acknowledgements

This research has been supported by NSF Grants CCF 1763702, AF 1901292, CNS 2148141, Tripods CCF 1934932, IFML CCF 2019844 and research gifts by Western Digital, WNCG IAP, UT Austin Machine Learning Lab (MLL), Cisco and the Archie Straiton Endowed Faculty Fellowship.

References

[1] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on information theory*, 56(4):1982–2001, 2010.

[2] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[3] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[4] M. Fehler and P. J. Keliher. *SEAM Phase I: Challenges of Subsalt Imaging in Tertiary Basins, with Emphasis on Deepwater Gulf of Mexico*. Society of Exploration Geophysicists, 01 2011.

[5] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir. Robust compressed sensing mri with deep generative priors. In *NeurIPS*, 2021.

[6] A. Jalal, S. Karmalkar, A. Dimakis, and E. Price. Instance-optimal compressed sensing via posterior sampling. In *ICML*, pages 4709–4720, 2021.

[7] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[8] G. Lin, F. Liu, A. Milan, C. Shen, and I. Reid. Refinenet: Multi-path refinement networks for dense prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[9] G. Lin, A. Milan, C. Shen, and I. Reid. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. In *CVPR*, July 2017.

[10] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman. Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development*, 12(3):1165–1187, 2019.

[11] F. Luporini, M. Louboutin, M. Lange, N. Kukreja, P. Witte, J. Hückelheim, C. Yount, P. H. J. Kelly, F. J. Herrmann, and G. J. Gorman. Architecture and performance of devito, a system for automated stencil computation. *ACM Trans. Math. Softw.*, 46(1), apr 2020.

[12] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[13] M. Mayall, L. Lonergan, A. Bowman, S. James, K. Mills, T. Primmer, D. Pope, L. Rogers, and R. Skeene. The response of turbidite slope channels to growth-induced seabed topography. *AAPG bulletin*, 94(7):1011–1030, 2010.

[14] H. Millán-Garrido. Geometry and kinematics of compressional growth structures and diapirs in the la popa basin of northeast mexico: Insights from sequential restoration of a regional cross section and three-dimensional analysis. *Tectonics*, 23(5), 2004.

[15] Y. Nakamura, S. Kodaira, S. Miura, C. Regalla, and N. Takahashi. High-resolution seismic imaging in the japan trench axis area off miyagi, northeastern japan. *Geophysical Research Letters*, 40(9):1713–1718, 2013.

[16] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

[17] Y. Song and S. Ermon. Improved techniques for training score-based generative models. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[18] P. Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

[19] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[20] H.-W. Zhou, H. Hu, Z. Zou, Y. Wo, and O. Youn. Reverse time migration: A prospect of seismic imaging methodology. *Earth-science reviews*, 179:207–227, 2018.