Learned imaging with constraints and uncertainty quantification

Felix J. Herrmann, Ali Siahkoohi, and Gabrio Rizzuti

{felix.herrmann,alisk,rizzuti.gabrio}@gatech.edu

School of Computational Science and Engineering, Georgia Tech

Seismic imaging: intro

Seismic imaging aims to recover **physical properties** of the Earth's interior based on **surface measurements**. Seismic sources (as airguns in a marine environment, or vibrator trucks on land) are placed at the surface, and seismic waves are excited, propagate through the Earth, and transmitted/reflected back to the surface, where data are recorded (by hydro- or geophones), yielding information about the subsurface.

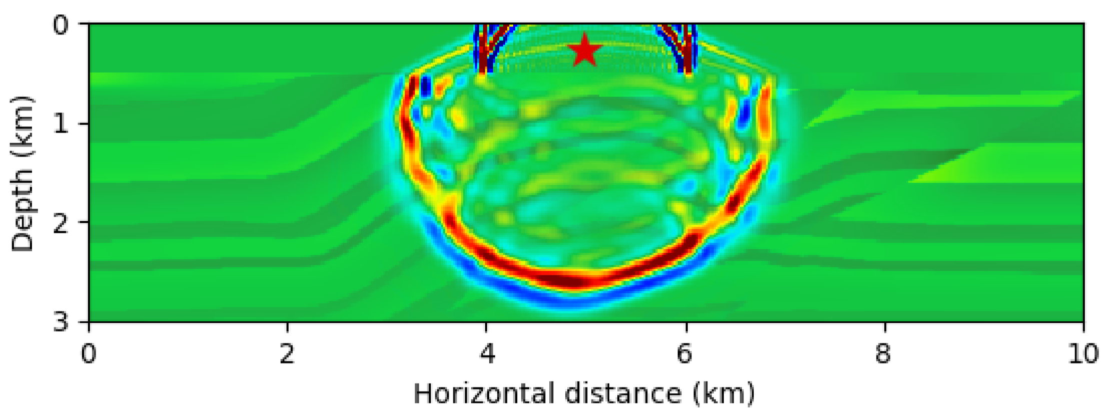

Figure 1: Wavefield snapshot overlayed to model

Deep priors, constraints, EM

We exploit the remarkable ease of **deep convolutional networks** to generate natural images, as added **implicit regularization** [4].

We consider the parameterized data model

$$\mathbf{d} = \mathcal{F}(\mathbf{q}, \mathbf{m}) + \boldsymbol{\varepsilon}, \qquad \mathbf{m} = G_\theta(\mathbf{z}) + \boldsymbol{\eta}, \tag{7}$$

s.t. $\mathbf{z} \sim N(0, \sigma_z^2 I)$, $p_\theta(\mathbf{m}\,|\,\mathbf{z}) \propto_{\mathbf{z},\theta} 1_C(\mathbf{m}) g_{\sigma^2_{\mathrm{m}}}(\mathbf{m} - G_\theta(\mathbf{z}))$ (where $g_{\sigma^2_{\mathrm{m}}}$ is Gaussian with variance σ^2_{m}), $\boldsymbol{\varepsilon} \perp (\mathbf{z}, \mathbf{q}, \mathbf{m})$, $\mathbf{q} \perp (\mathbf{z}, \mathbf{m})$. We end up with the **data model**

$$\begin{aligned} p_\theta(\mathbf{q}, \mathbf{d}) &= \iint p_\theta(\mathbf{q}, \mathbf{d}, \mathbf{z}, \mathbf{m})\, \mathrm{d}\mathbf{m}\, \mathrm{d}\mathbf{z}, \\ &= \iint p(\mathbf{q}, \mathbf{d}\,|\,\mathbf{m}) p_\theta(\mathbf{m}\,|\,\mathbf{z}) p(\mathbf{z})\, \mathrm{d}\mathbf{m}\, \mathrm{d}\mathbf{z}, \end{aligned} \tag{8}$$

and maximum-likelihood estimation problem

$$\min_\theta \mathbb{E}_{(\mathbf{q},\mathbf{d}) \sim p_{\mathrm{data}}(\mathbf{q},\mathbf{d})} - \log p_\theta(\mathbf{q}, \mathbf{d}). \tag{9}$$

Note that the **latent variables** $(\mathbf{z}, \mathbf{m})$ are **jointly distributed** (coupled) with data $(\mathbf{q}, \mathbf{d})$.

The **expectation maximization** method (EM) is based on the identity:

$$\partial_\theta \log p_\theta(\mathbf{q}, \mathbf{d}) = \mathbb{E}_{(\mathbf{z},\mathbf{m}) \sim p_\theta(\mathbf{z},\mathbf{m}\,|\,\mathbf{q},\mathbf{d})} \partial_\theta \log p_\theta(\mathbf{q}, \mathbf{d}, \mathbf{z}, \mathbf{m}). \tag{10}$$

Having set the loss $\mathcal{L}_\theta = -\log p_\theta(\mathbf{q}, \mathbf{d}, \mathbf{z}, \mathbf{m})$,

$$\begin{aligned} \mathcal{L}_\theta(\mathbf{q}, \mathbf{d}\,;\mathbf{z}, \mathbf{m}) &= \frac{1}{2\sigma_{\mathrm{d}}^2} \|\mathbf{d} - \mathcal{F}(\mathbf{q}, \mathbf{m})\|^2 \\ &+ \frac{1}{2\sigma_{\mathrm{m}}^2} \|\mathbf{m} - G_\theta(\mathbf{z})\|^2 + \frac{1}{2\sigma_{\mathrm{d}}^2} \|\mathbf{z}\|^2, \end{aligned} \tag{11}$$

we alternate the following steps:

- (E) update $\mathbf{m}$ via (6) applied to (11) ($\mathbf{z}$ fixed), and sample $\mathbf{z} \sim p_\theta(\mathbf{z}\,|\,\mathbf{m})$ with Langevin dynamics ($\mathbf{m}$'s fixed);
- (M) update $\theta \leftarrow \theta - t\nabla_\theta \sum_{\mathbf{z},\mathbf{m}} \|\mathbf{m} - G_\theta(\mathbf{z})\|^2$ (not accounting for the dependency of the sampled $\mathbf{z}, \mathbf{m}$ wrt θ, according to (10)).

Classical setting

The observed data are pairs of seismic point source/recorded waveforms $(\mathbf{q}, \mathbf{d}) \sim p_{\mathrm{data}}(\mathbf{q}, \mathbf{d})$. The **data likelihood** $p(\mathbf{q}, \mathbf{d}\,|\,\mathbf{m})$, given the physical parameters $\mathbf{m}$ (also, $\mathbf{q} \perp \mathbf{m}$), is:

$$\mathbf{d} = \mathcal{F}(\mathbf{q}, \mathbf{m}) + \boldsymbol{\varepsilon}, \tag{1}$$

where the **forward map** $\mathcal{F}$ is defined by

$$\begin{aligned} \mathcal{F}(\mathbf{q}, \mathbf{m}) &= RA(\mathbf{m})^{-1}\mathbf{q}, \\ A(\mathbf{m}) &= \mathbf{m}\partial_{tt} - \Delta. \end{aligned} \tag{2}$$

$A(\mathbf{m})$ is the **wave equation** system, R is a restriction-to-receiver operator, and $\boldsymbol{\varepsilon}$ is noise ($\boldsymbol{\varepsilon} \perp (\mathbf{q}, \mathbf{m})$).

Due to **numerical stability** of the forward problem, we must impose **hard constraints** $\mathbf{m} \in C$ (i.e. $\mathbf{m}$ is uniformly distributed on C), for a convex set C which might comprise box constraints, total variation norm bounds, etc [1].

When $\boldsymbol{\varepsilon} \sim N(0, \sigma_{\mathrm{d}}^2 I)$ is Gaussian, we end up solving the **nonlinear least-squares** problem:

$$\begin{aligned} &\min_{\mathbf{m} \in C} \mathbb{E}_{(\mathbf{q},\mathbf{d}) \sim p_{\mathrm{data}}(\mathbf{q},\mathbf{d})} \mathcal{L}(\mathbf{q}, \mathbf{d}\,;\mathbf{m}), \\ &\mathcal{L}(\mathbf{q}, \mathbf{d}\,;\mathbf{m}) = \frac{1}{2\sigma_{\mathrm{d}}^2} \|\mathbf{d} - \mathcal{F}(\mathbf{q}, \mathbf{m})\|^2. \end{aligned} \tag{3}$$

For 3D problems, $\mathbf{m} \in \mathbb{R}^{n^3}$ ($n \approx 1000$), $\mathbf{d} \in \mathbb{R}^{n^3}$, and the typical data sample $\{(\mathbf{q}_i, \mathbf{d}_i)\}_{i=1}^N$ size is $N = O(n)$. The forward map $\mathcal{F}$ is very **expensive to evaluate** and gradient descent methods tend to **converge slowly**.

Linearization and CS

By **linearizing** the problem around a known (kinematically correct) background model $\mathbf{m}_{\mathrm{bg}}$

$$\mathcal{F}(\mathbf{q}, \mathbf{m}_{\mathrm{bg}} + \Delta\mathbf{m}) \approx \mathcal{F}(\mathbf{q}, \mathbf{m}_{\mathrm{bg}}) + J(\mathbf{q}, \mathbf{m}_{\mathrm{bg}})\Delta\mathbf{m}, \tag{4}$$

we can formally treat (3) as a linear problem.

Motivated by **compressive sensing** (CS), we consider simultaneous-source experiments

$$\mathbf{q}(\mathbf{w}) = \textstyle\sum_i w_i \mathbf{q}_i, \quad \mathbf{d}(\mathbf{w}) = \sum_i w_i \mathbf{d}_i, \tag{5}$$

where $\mathbf{w} \sim N(0, I)$.

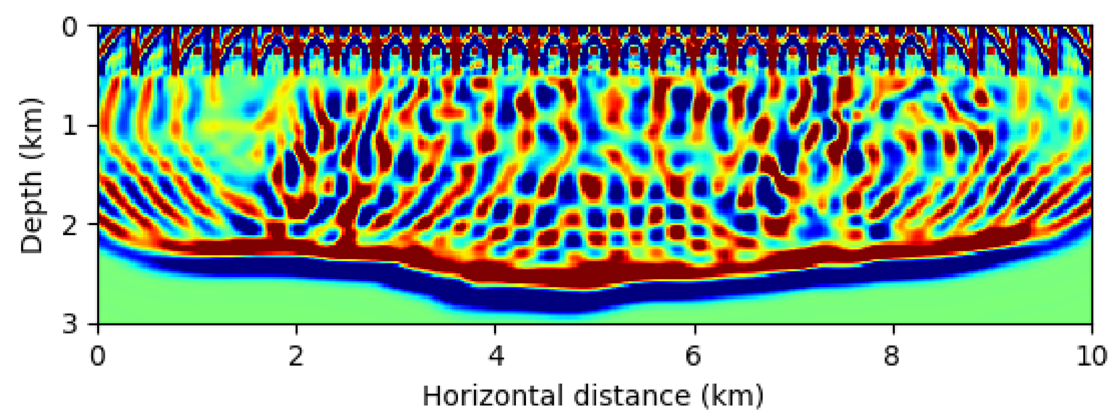

Figure 8: Simultaneous-source wavefield

We employ a **linearized Bregman** [3] algorithm with small-batch approximations of the loss gradient [2]:

$$\begin{aligned} \mathbf{g}_k &\approx \nabla_{\mathbf{m}_k} \mathbb{E}_{(\mathbf{q},\mathbf{d}) \sim p(\mathbf{q},\mathbf{d})}\, \mathcal{L}(\mathbf{q}, \mathbf{d}\,;\mathbf{m}), \\ \tilde{\mathbf{m}}_{k+1} &= \mathbf{m}_k - t_k \mathbf{g}_k, \\ \mathbf{m}_{k+1} &= P_C(\tilde{\mathbf{m}}_{k+1}), \end{aligned} \tag{6}$$

with dynamic steplength t_k. P_C is a projection on the constraint set C.

Results

Inversion results

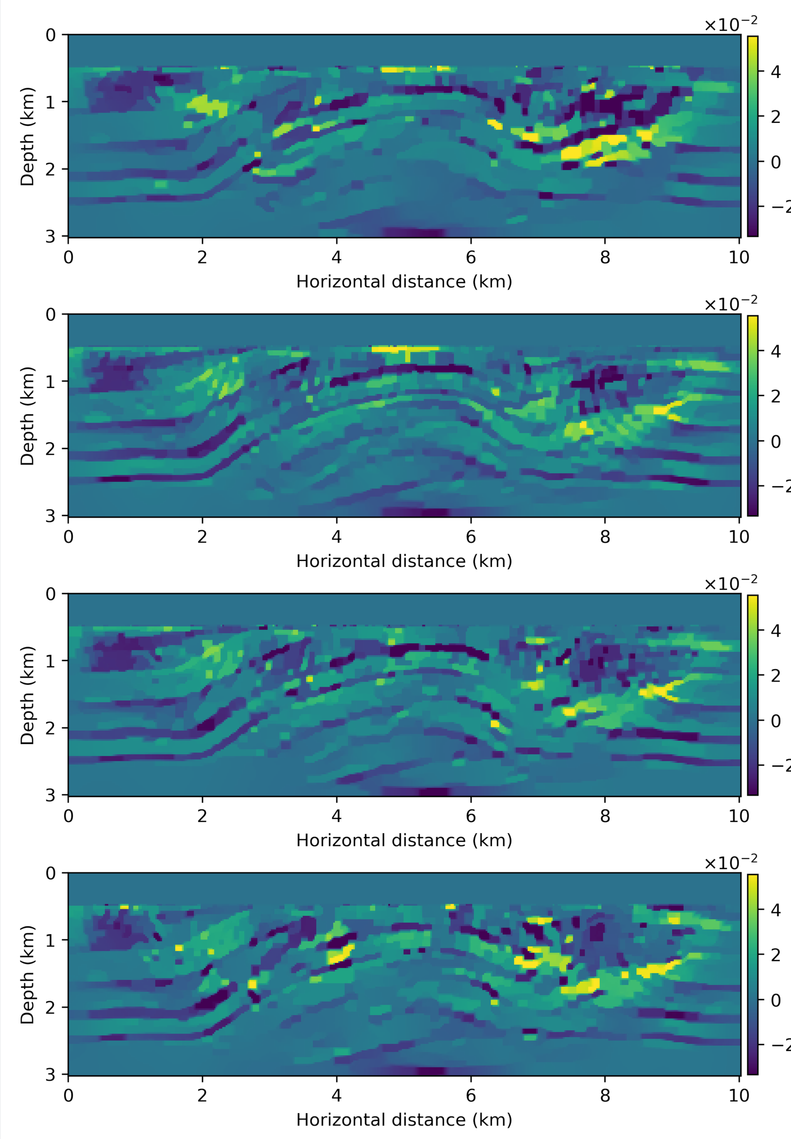

Figure 2: $\mathbf{m}$ (for different $\mathbf{q}, \mathbf{d}$'s)

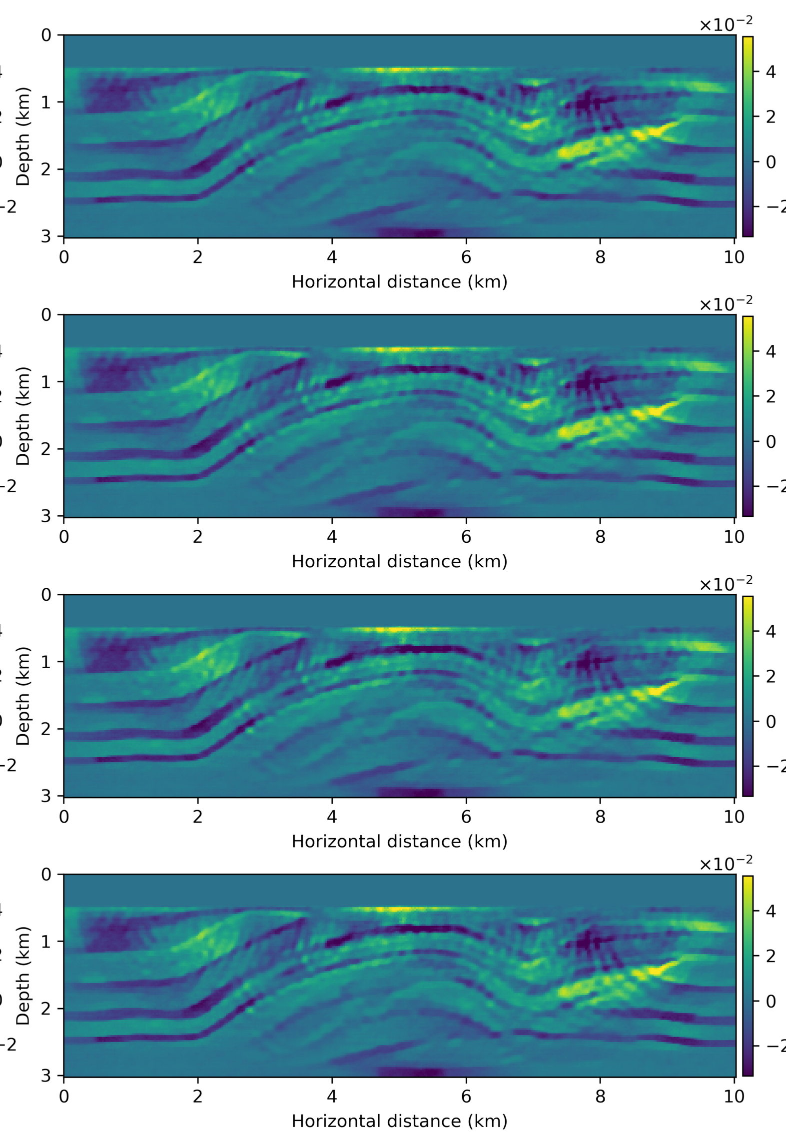

Figure 3: $G_\theta(\mathbf{z})$ (for different $\mathbf{q}, \mathbf{d}$'s)

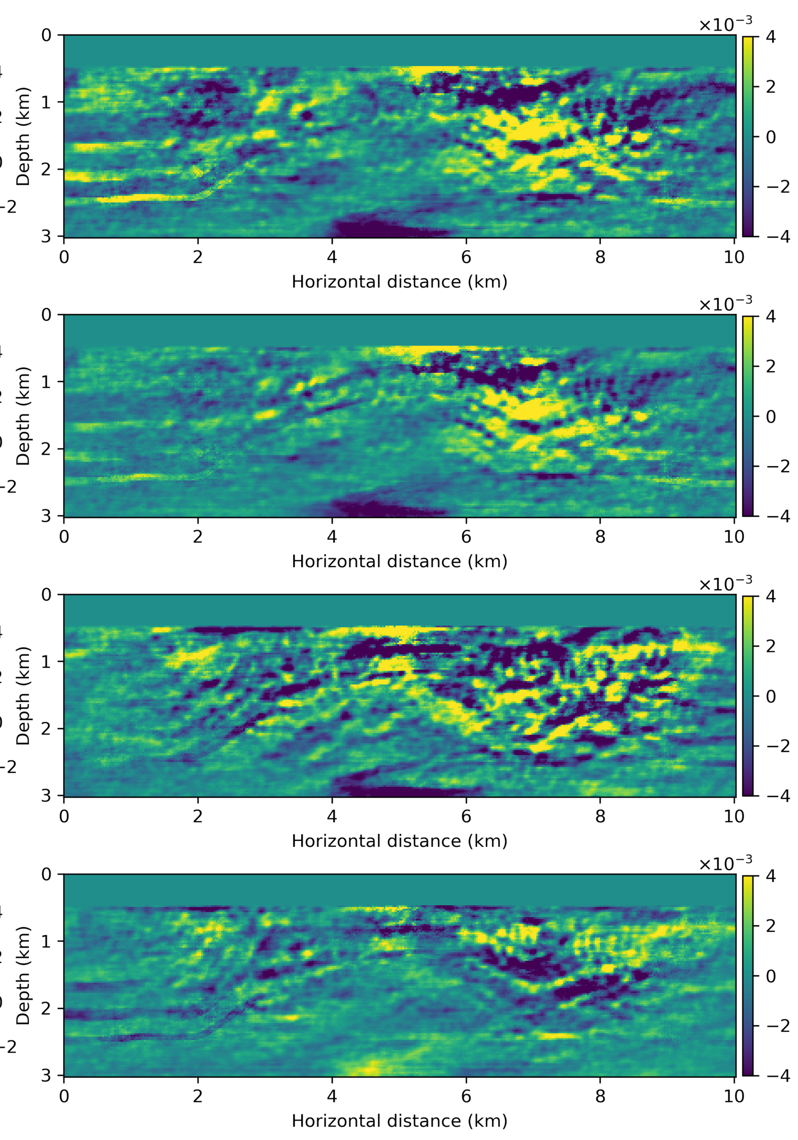

Figure 4: $G_\theta(\mathbf{z}) - G_\theta(\mathbf{z_0})$

Generator pointwise statistics

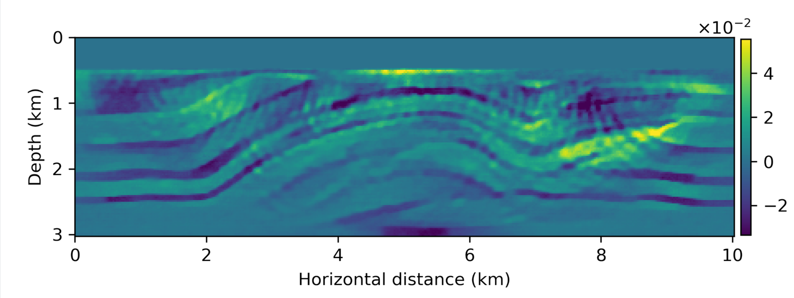

Figure 5: Mean of $G_\theta(\mathbf{z})$

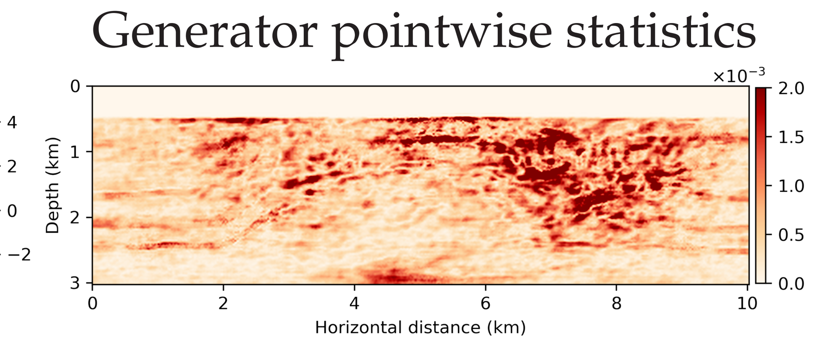

Figure 6: Standard deviation of $G_\theta(\mathbf{z})$

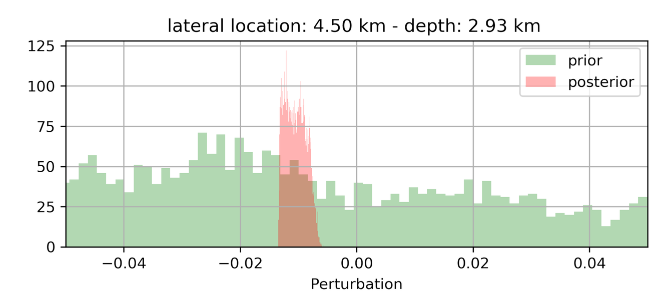

Figure 7: Grid point histogram

References

[1] P. Bas, R. S. Brendan, and F. J. Herrmann. Projection methods and applications for seismic nonlinear inverse problems with multiple constraints. *Geophysics*, 84(2):R251–R269, 2019.

[2] P. A. Witte, M. Louboutin, F. Luporini, G. J. Gorman, and F. J. Herrmann. Compressive least-squares migration with on-the-fly Fourier transforms. *Geophysics*, 84(5):1–76, 2019.

[3] D. A. Lorenz, F. Schöpfer, and S. Wenger. The linearized Bregman method via split feasibility problems: analysis and generalizations. *SIAM Journal on Imaging Sciences*, 7(2):1237–1262, 2014.

[4] S. Dittmer, T. Kluth, P. Maass, and D. Otero Baguer. Regularization by Architecture: A Deep Prior Approach for Inverse Problems. *Journal of Mathematical Imaging and Vision*, 2019.

[5] T. Han, Y. Lu, S.-C. Zhu, and Y. N. Wu. Alternating back-propagation for generator network. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[6] S. Zhang, A. E. Choromanska, and Y. LeCun. Deep learning with elastic averaging SGD. In *Advances in Neural Information Processing Systems*, pages 685–693, 2015.

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