# Multi-Dimensional Deconvolution with stochastic gradient descent

## Introduction

Multi-Dimensional Deconvolution (MDD) is a popular technique in the seismic processing and imaging domains used to suppress overburden effects by deconvolving the up- and down-going components of the recorded seismic data at a datum of interest. Although the original theory of MDD dates back to the seminal work of Amundsen (2001), for almost two decades the geophysical community has only embraced a 1D approximation of such a theory as it leads to an efficient element-wise, stabilized division in the frequency-wavenumber or Radon domains. Despite the theoretical superiority, the arguments that MDD is severely ill-posed and difficult to stabilize alongside its extreme computational cost have for long time hindered its widespread adoption in industrial settings. Recently, Vargas et al. (2021a) showed that a time-domain formulation of the problem in combination with physics-based preconditioners (i.e., causality, reciprocity, frequency-wavenumber locality) can lead to a stable deconvolution process.

In this work, after reinterpreting the MDD cost function as a finite-sum functional, we solve the associated inverse problem using stochastic gradient descent algorithms. This is primarily motivated by the need to reduce the overall computational cost of MDD and supported by the physical argument that nearby sources are likely to provide redundant contributions to the gradients, which drive the time-domain MDD inversion. The proposed algorithm is successfully applied to a synthetic and a field dataset.

## Theory

The up- ($p^-(\mathbf{x}_{VS}, \mathbf{x}_S, \omega)$) and down-going ($p^+(\mathbf{x}_R, \mathbf{x}_S, \omega)$) components of the seismic wavefield from a source $\mathbf{x}_S$ to a line of receivers $\mathbf{x}_R$ at a given datum $\partial\mathbb{D}$ and another receiver $\mathbf{x}_{VS}$ placed at any location below $\partial\mathbb{D}$ are linked to the local reflection response $R(\mathbf{x}_R, \mathbf{x}_{VS}, \omega)$ via the following integral equation in the frequency domain (Amundsen, 2001):

$$p^-(\mathbf{x}_{VS}, \mathbf{x}_S, \omega) = \int_{\partial\mathbb{D}} p^+(\mathbf{x}_R, \mathbf{x}_S, \omega) R(\mathbf{x}_R, \mathbf{x}_{VS}, \omega) d\mathbf{x}_R \tag{1}$$

Whilst the forward problem is well-posed for a single source, the associated inverse problem is severely ill-posed. It therefore requires availability of multiple sources, which ideally should be equal to or exceeds the number of receivers along the datum $\partial\mathbb{D}$ (i.e., $N_s \geq N_r$). A time-domain formulation of MDD, as presented in Luiken and van Leeuwen (2020), can be written as follows:

$$\mathbf{r}_t = \underset{\mathbf{r}_t \in \mathbb{C}}{\operatorname{argmin}} \frac{1}{2} ||\mathbf{p}_t^- - \mathbf{P}_t^+ \mathbf{r}_t||_2^2 \tag{2}$$

where $\mathbf{p}_t^-$ is a vector of size $[N_s N_{vs} N_t \times 1]$ and $\mathbf{r}$ is a vector of size $[N_r N_{vs} N_t \times 1]$ obtained by stacking traces for all available pairs of sources (or receivers) and virtual sources. $\mathbf{P}_t^+$ is a linear operator applying multi-dimensional convolution in the time domain as detailed in Ravasi and Vasconcelos (2021). The solution is further enforced to belong to a sub-subspace of interest, $\mathbf{r}_t \in \mathbb{C}$, based on physical constraints that the wavefield is expected to satisfy (i.e., reciprocity and/or causality). Note that whilst the causality constraint can be enforced when inverting for a single virtual source at the time, the reciprocity constraint requires dealing with many virtual sources at once, which must be co-located with receivers. In the following, we will refer to the solution of equation 2 by means of LSQR as full-gradient MDD.

As the forward problem in equation 1 is valid for any of the available sources, the functional in equation 2 can be equivalently written as a finite-sum functional (Figure 1a):

$$\mathbf{r}_t = \underset{\mathbf{r}_t \in C}{\operatorname{argmin}} \frac{1}{2} \sum_{i_s=1}^{N_s} ||\mathbf{p}_{t,i_s}^- - \mathbf{P}_{t,i_s}^+ \mathbf{r}_t||_2^2 \tag{3}$$

where $\mathbf{p}_{t,s_i}^-$ is a vector of size $[N_{vs} N_t \times 1]$ that contains the up-going wavefield originated from the i-th source. Similarly, $\mathbf{P}_{t,i_s}^+$ is a linear operator performing MDC with the down-going wavefield of the i-th
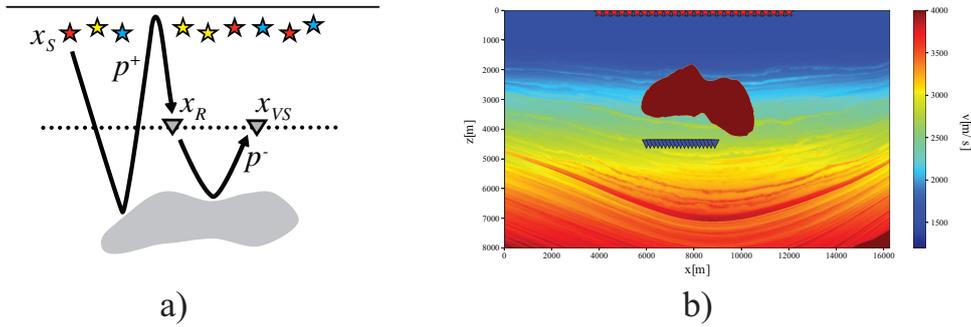
Figure 1: a) Schematic representation of the stochastic MDD algorithm. Red, yellow, and cyan sources represent three different batches. Grey triangles indicate a receiver and a virtual source along a sample raypath composed of the down-going ($p^+$) and up-going ($p^-$) components linked together via equation 1. b) Synthetic subsalt velocity model. Red stars and blue triangles indicate sources and receivers.

source. Note that a physical preconditioner may be added to this functional similar to that in equation 2. Drawing upon the vast literature of stochastic gradient algorithms (e.g., Saad, 1998), the inverse problem in equation 3 can be conveniently solved in a mini-batch fashion. More specifically, rather than computing the exact gradient of equation 3 at each iteration and using it to update the current $\mathbf{r}_t$ vector, the available sources are grouped in batches $S_{batch}$ of size $N_{s,batch} < N_s$ such that the gradient of each batch $S_{batch}$ and the update of the model vector at step k are computed as follows:

$$\mathbf{g}_{t,S_{batch}} = - \sum_{i_s \in S_{batch}} \mathbf{P}_{t,i_s}^{+H}(\mathbf{p}_{t,i_s}^- - \mathbf{P}_{t,i_s}^+ \mathbf{r}_t), \qquad \mathbf{r}_t^k = \mathbf{r}_t^{k-1} - \alpha \mathbf{g}_{t,S_{batch}}. \tag{4}$$

The step-size is chosen using a semi-heuristic procedure based on the Landweber iteration, $\alpha < \sigma(\mathbf{P}^{+H}\mathbf{P}^+)$ where $\sigma(\mathbf{P}^{+H}\mathbf{P}^+)$ is equivalent to the maximum squared singular value of the down-going wavefield in the frequency domain (Luiken and van Leeuwen, 2020). Nesterov momentum is also used to further accelerate convergence. In the following this approach is called the N-SGD algorithm. The word *epoch* will be used when referring to a set of gradient steps that utilize all of the available sources in a dataset. Both algorithms are implemented using the PyLops framework (Ravasi and Vasconcelos, 2021).

**Numerical results**

The first example is based on the synthetic model presented in Vargas et al. (2021b) (Figure 1b), where MDD is performed on the up- and down-going wavefields at a depth of $z_R = 4500m$. The acquisition geometry comprises of 201 sources 40m apart and 151 receiver with spacing of 20m. Such wavefields are estimated from surface seismic data by means of a Marchenko-based redatuming process that inevitably introduces coherent artefacts; we therefore expect MDD to be highly unstable and physical preconditioning are required to produce satisfactory results when working with full-gradient MDD (Vargas et al., 2021a). The ideal reflection response modelled in the truncated medium is shown in Figure 2a. The response retrieved by means of full-gradient MDD using a single virtual-source is clearly contaminated by strong dipping events (Figure 2b). This result can be dramatically improved by solving for multiple virtual sources at once and introducing a reciprocity preconditioner (Figure 2c). On the other hand, the reflection responses produced by means of single-virtual source stochastic MDD using a batch size of 32 sources is much cleaner than its full-gradient counterpart without the requirement to introduce physical preconditioners (Figure 2d). Although not shown here, similar results are achieved for multi-virtual-source stochastic MDD. However, whilst the use of a preconditioner seems to be critical for the former method, the improvement is fairly minor in the latter case making it possible to work with a single or small group of virtual sources at the same time without seriously compromising the quality of the reconstruction. This finding is of importance for the application of MDD to large-scale, three dimensional datasets where solving for the entire set of virtual sources at the same time may be beyond reach of our current compute capabilities (Ravasi and Vasconcelos, 2021). Looking at the norms of the model error (Figure 2e), we observe how those of the full-gradient MDD algorithms reduce in the first few epochs, and start to grow as artefacts become more prominent in later epochs. Whilst this result highlights the
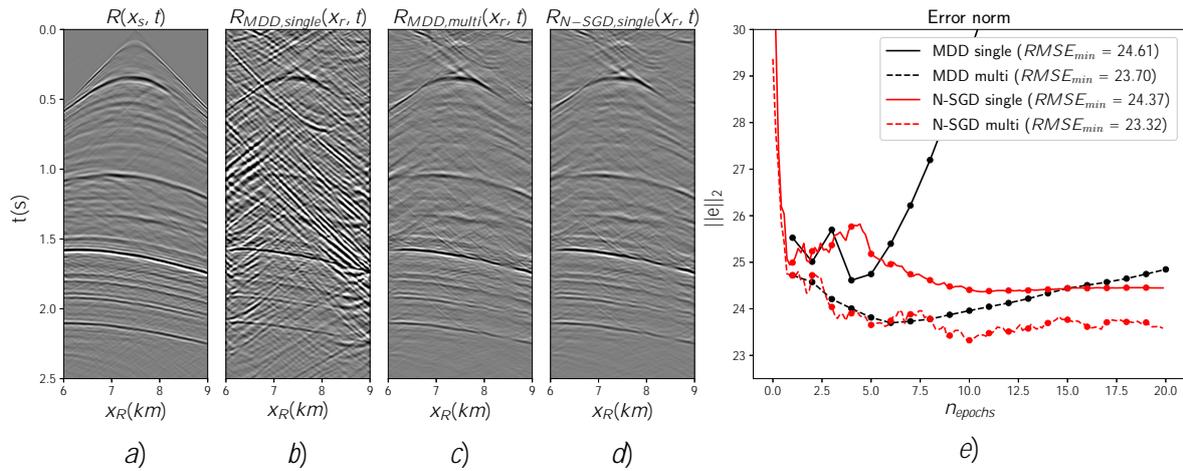
Figure 2: a) Ideal reflection response for the subsalt model generated without overburden effects. b) Single-virtual-source, full-gradient MDD, c) multi-virtual-source, full-gradient MDD, d) single-virtual-source, stochastic MDD, and e) error norms as function of epochs for the different solvers. Dots display the values of the respective norms at the end of each epoch.

importance of early stopping in full-gradient MDD, in practical applications the error norm cannot be computed; as no indication of data overfitting can be observed by looking at a proxy norm such as the residual norm, it becomes difficult to objectively and automatically decide the maximum number of epochs for MDD. A more stable convergence is instead observed for the stochastic MDD algorithms.

Moving onto to the field dataset, a 2D line composed of 120 sources and 180 receivers with 50m and 25m spacing, respectively, is selected from the 3D OBC Volve dataset acquired by Statoil in 2010. Pressure and vertical particle velocity components are pre-processed using a similar flow to that described in Ravasi et al. (2015) and used as input to full-gradient MDD and stochastic MDD (with batch size of 32 sources). Here, both causality and reciprocity preconditioners are used to aid the solution of the different MDD algorithms. Figure 3 displays the estimated reflection responses by means of cross-correlation and MDD. Note that whilst the different MDD algorithms are all able to some extent to remove the cross-talk artefacts visible in the cross-correlation gather, the single- and multi-virtual-source stochastic MDD algorithms produce cleaner wavefield reconstructions compared to their full-gradient counterparts (see, for example, the event indicated by red arrows). Finally, to better appreciate the impact of MDD and the difference between the full-gradient and stochastic algorithms, the full and up-going pressure data as well as the estimated reflection responses from the multi-virtual-source MDD algorithms are imaged by means of reverse-time migration (Figure 4). We can clearly observe that both the receiver ghost and higher-order free-surface multiples are successfully suppressed in the two MDD images. Moreover, the image created from the reflection response estimated by means of stochastic MDD is visibly cleaner, especially in the shallow part of the subsurface.

**Conclusions**

We have presented a novel formulation of Multi-Dimensional Deconvolution as a finite-sum problem. This allows leveraging stochastic gradient descent algorithms, where gradients at each step are computed using a (small) random subset of sources. Synthetic and field data examples validate the effectiveness of the proposed methodology, and its superiority over full-gradient MDD both in terms of convergence speed and quality of the estimated local reflection response.
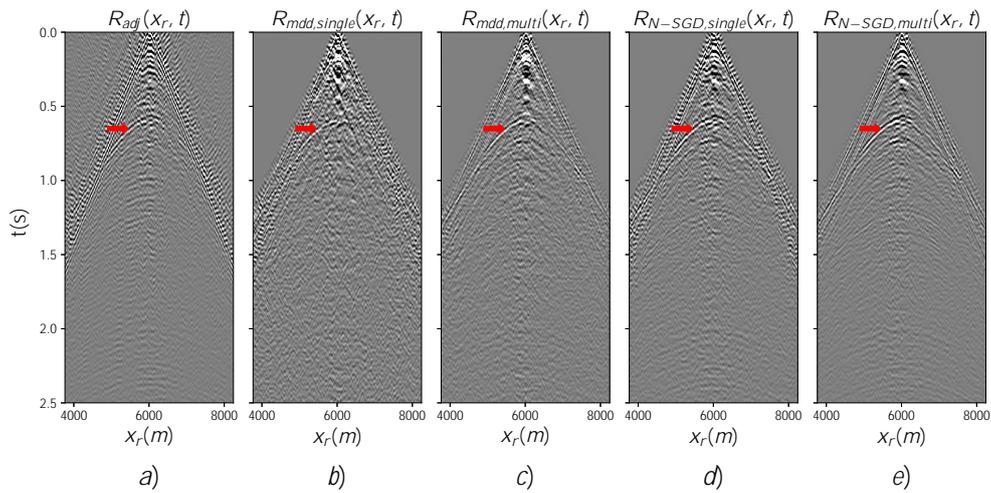
**Acknowledgements**

Figure 3: Reflection responses estimated from the Volve field dataset using a) cross-correlation b) single-virtual-source full-gradient MDD, c) multi-virtual-source full-gradient MDD, d) single-virtual-source N-SGD MDD, and e) multi-virtual-source N-SGD MDD.
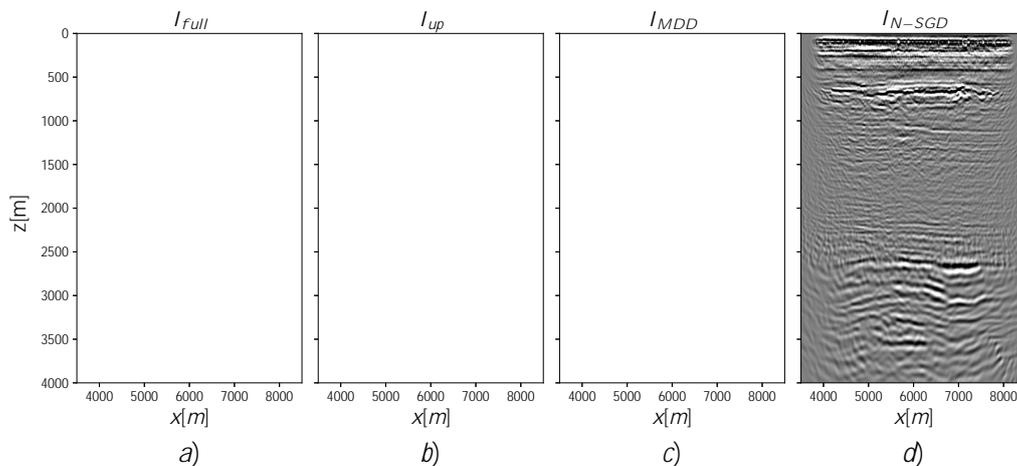


Figure 4: Images obtained using a) full data, b) up-going data, c) reflection response from full-gradient MDD, and d) reflection response from stochastic MDD.

# References

[1] Amundsen, L. [2001] Elimination of free-surface related multiples without need of a source wavelet. Geophysics, 66, 327–341.

[2] Luiken, N., and van Leeuwen, T. [2020] Seismic wavefield redatuming with regularized multidimensional deconvolution. Inverse Problems, 36, 095010.

[3] Ravasi, M. and Vasconcelos, I. [2020] PyLops - A linear-operator Python library for scalable algebra and optimization. SoftwareX 11, 100361.

[4] Ravasi, M., and Vasconcelos, I. [2021] An open-source framework for the implementation of large-scale integral operators with flexible, modern hpc solutions - Enabling 3D Marchenko imaging by least-squares inversion. Geophysics, 86, WC177–WC194.

[5] Ravasi, M., Vasconcelos, I., Curtis, A., and Kritski, A. [2015] Multi-dimensional free-surface multiple elimination and source deblending of Volve OBC data. 77th Conference and Exhibition, EAGE, Extended Abstracts.

[6] Saad, D. [1998] On-line learning in neural networks. Cambridge University Press.

[7] Vargas, D., Vasconcelos, I., Ravasi, M., and Luiken, N. [2021a] Time-domain multidimensional deconvolution: A physically reliable and stable preconditioned implementation. Remote Sensing, 13, 3683.

[8] Vargas, D., Vasconcelos, I., Ravasi, M., and Sripanich, Y. [2021b] Scattering-based focusing for imaging in highly-complex media from band-limited, multi-component data. Geophysics, 86, 1–64.