Stochastic Multi-Dimensional Deconvolution
Matteo Ravasi\textsuperscript{D}, Tamin Selvan\textsuperscript{D}, and Nick Luiken

Abstract—Geophysical measurements such as seismic datasets contain valuable information that originate from areas of interest in the subsurface; these seismic reflections are, however, inevitably contaminated by other events created by waves reverberating in the overburden. Multidimensional deconvolution (MDD) is a powerful technique used at various stages of the seismic processing sequence to create ideal datasets deprived of such overburden effects. While the underlying forward problem holds for a single source, a successful inversion of the MDD equations requires the availability of a large number of sources alongside prior information, possibly introduced in the form of physical constraints (e.g., reciprocity and causality). In this work, we present a novel formulation of time-domain MDD based on a finite-sum functional. The associated inverse problem is then solved by means of stochastic gradient descent algorithms, where the gradients at each iteration are computed using a small subset of randomly selected sources. Through synthetic and field data examples, we show that the proposed method converges more stably than the conventional approach based on full gradients. Stochastic MDD represents a novel, efficient, and robust strategy to deconvolve seismic wavefields in a multidimensional fashion.

Index Terms—Deconvolution, inverse problems, seismic, stochastic gradient.

I. INTRODUCTION

MULTIDIMENSIONAL deconvolution (MDD) is a versatile technique used in seismic processing and imaging to suppress the effect of an unwanted overburden; this is simply achieved by deconvolving the up- and down-going components of the recorded seismic data at a certain datum of interest. In this context, the word overburden may simply refer to the free surface when data are acquired via a streamer or ocean-bottom cable (OBC) acquisition system, or a portion of the earth when recordings are available at a certain depth within the subsurface in either a horizontal borehole or estimated by means of model-based (e.g., [1]–[3]) or data-driven (e.g., [4]–[6]) redatuming techniques. More recently, the applicability of MDD has also been extended to laboratory acoustic measurements with the aim of removing from the recorded data unwanted reflections originated at the boundaries of the experimental setup [7]. Such a finding may also have broader implications in other domains of science such as nondestructive testing or medical imaging; in both cases, acoustic waves originated from discontinuities inside a probed body are in fact polluted by strong reverberations from the rigid boundary between such a body and the fluid (i.e., air or water) that encompasses it.

The original theory of MDD dates back to the seminal work of [8] and it has been later formalized within the context of seismic interferometry [9], [10]. However, for a long time, the geophysical community has only embraced a 1-D approximation of such a theory as it leads to an efficient element-wise, stabilized division in the frequency–wavenumber (\(f–k\)) or linear Radon (\(r–p\)) domains [11], [12], sometimes referred to as up/down deconvolution (UDD). While such an approximation is valid for a horizontally invariant overburden, its accuracy deteriorates significantly when this assumption is not met. A special case is represented by a dipping seabed for the free-surface multiple attenuation scenario, as the only interface in the overburden is represented by the sea–air interface, which is not aligned with the seafloor. Within this context, Boiero and Bagaini [13] have recently studied the validity of the UDD method against the more accurate MDD method and concluded that the latter must be used for seabed dips beyond 1°. Similarly, when the purpose of MDD is that of suppressing the effect of a highly complex overburden within an imaging context, the 1-D approximation is never valid. This explains the inability of single-channel deconvolution between the source (down-going) and receiver (up-going) wavefields to handle crosstalk artifacts when jointly imaging primaries and multiples [14], [15].

Despite the theoretical superiority of MDD over UDD, the arguments that MDD is a severely ill-posed and difficult to stabilize inverse problem alongside its extreme computational cost have for a long time hindered the widespread adoption of such an algorithm in industrial settings. Early attempts to mitigate the ill-posed nature of such an inverse problem have been reported by Minato et al. [16] that employed singular value decomposition (SVD) when solving MDD in the frequency domain. More recently, van der Neut and Herrmann [17] proposed a time-domain formulation of the problem to naturally regularize the inverse problem and used sparsity-promoting inversion to further mitigate the effect of noise in the input data. Along similar lines, van der Neut et al. [18] and [19] proposed to introduce a preconditioner to enforce reciprocity in the coveted target response. Vargas et al. [20] combined three physics-based preconditioners (i.e., causality, reciprocity, and frequency–wavenumber locality) alongside providing a comprehensive analysis of their individual and combined impact on the solution of MDD in the presence of noisy input data retrieved by means of the so-called Scattering-Rayleigh Marchenko redatuming method (SRM— [21]). Although these studies have shown that the time-domain approach is undoubt-
edly superior to its frequency-domain counterpart, one major issue when solving ill-posed inverse problems with iterative solvers is represented by the so-called semi-convergence behavior. Semi-convergence is the observed phenomenon in that the error norm starts to increase after a certain number of iterations, whereas the residual norm keeps monotonically decreasing. However, since in practical applications we can only access the residual (and not the error), we are oblivious to this effect. A number of attempts have been made to identify and mitigate semi-convergence in practice [22, Chapter 6], however, a thorough theoretical understanding of semi-convergence does not exist at the moment. This phenomenon simply arises because iterative solvers resolve components in eigenvector directions corresponding to large eigenvalues first, and components associated with the smaller eigenvalues later (see [22, Chapter 6], [23, p. 89]). For ill-posed problems, small eigenvalues are usually below the noise level and should not be resolved; by continuing the iterations and resolving components corresponding to small eigenvalues, we allow noise to enter the solution. Because of this behavior, the number of iterations plays the role of the regularization parameter for iterative solvers. Estimating a suitable number of iterations upfront is a notoriously difficult problem. Although a number of methods have been proposed, none of them have shown to be either applicable to or successful for the MDD problem [19]. Therefore, methods that are far less sensitive to this issue are extremely needed in practice.

In this work, we first reinterpret the cost function associated with time-domain MDD as a finite-sum functional. We then propose to solve the associated inverse problem using stochastic gradient descent algorithms. This is primarily motivated by the need to reduce the overall computational cost of gradient computations in MDD and supported by the physical argument that sources geographically close to each other are likely to provide redundant contributions to the gradients that drive the time-domain MDD inversion. As a by-product of this new algorithmic choice, we observe overall increased stability in the convergence properties of MDD, making it less reliant on the choice of the maximum number of iterations or stopping criterion for the gradient-based solver of choice. Moreover, we consistently observe faster (or at least not slower) convergence when using stochastic gradients instead of full gradients. Similar behavior has been previously reported in the literature [24], [25], where authors have shown that minimizing strongly convex finite sums is provably faster in expectation than is possible when ignoring the finite sum structure of the problem. The proposed algorithm is applied to three synthetic datasets and the Volve field dataset. As far as synthetic data are concerned, the first and third examples focus on the demultiple problem of up- and down-going data from an OBC acquisition geometry along a nonflat seabed. The second example considers instead subsurface-to-surface wavefields obtained by means of SRM redatuming, which requires us to deal with a severely ill-posed MDD problem [20]. This article is structured as follows. First, the theory of stochastic MDD is presented. This is followed by a section containing a set of numerical examples. We conclude by discussing the benefits of the proposed methodology and its outstanding challenges.

II. THEORY

The up- \((p^- (x_{VS}, x_S, f))\) and down-going \((p^+ (x_R, x_S, f))\) components of the seismic wavefield from a source \(x_S\) to a line of receivers \(x_R\) at a given datum \(\partial D\) and another receiver \(x_{VS}\) at any location below \(\partial D\) are linked to the local reflection response \(R(x_R, x_{VS}, f)\) via the following multidimensional convolution (MDC) integral in the frequency domain [8], [9]:

\[
p^-(x_{VS}, x_S, f) = \int_{\partial D} p^+(x_R, x_S, f) R(x_R, x_{VS}, f) dx_R \tag{1}
\]

where \(f\) is the frequency. In this context, the local reflection response is intended as the seismic response of an ideal medium that is equivalent to the original medium below the datum \(\partial D\) and homogeneous above, and therefore deprived of any interaction with the overburden. The process of inverting (1) is usually referred to in the literature as MDD.

From a mathematical point of view, while the forward problem holds for a single source, the solution of (1) requires availability of multiple sources, which ideally should be equal to or exceed the number of receivers along the datum \(\partial D\) (i.e., \(N_r \geq N_e\)). Moreover, despite (1) calls for an infinitely extended boundary \(\partial D\), in practice, this cannot be achieved. By truncating the contributions of the spatial integral to the (finite) extent of the receiver array, the modeled wavefield in the right-hand side of (1) does inevitably differ from the recorded up-going wavefield on the left-hand side of the same equation. As a consequence of this, MDD is a severely ill-posed inverse problem (see, e.g. [16]), where the severity of such ill-conditioning depends on various factors including the receiver array aperture, the depth of the receiver array, distance from the source, the complexity of the overburden, and the frequency content in the data. Finally, it is also common practice to consider multiple virtual sources \(x_{VS}\) placed along the same datum \(\partial D\) and express (1) in a compact matrix–matrix notation for each frequency \(f\)

\[
P_f = P_f R_f \tag{2}
\]
where $P^+_f$ is a matrix of size $[N_t \times N_{vs}]$, $P^+_f$ is a matrix of size $[N_t \times N_t]$, and $R_f$ is a matrix of size $[N_t \times N_t]$. This system of equations can be solved with direct methods [9], [16] or iterative Block–Krylov solvers [26]. Alternatively, a time-domain formulation can be written as [20]

$$r_t = \arg\min_{r_t \in \mathcal{A}} \frac{1}{2} \| p_t^+ - P^+_f r_t \|_2^2$$

(3)

where $p_t^+$ is a vector of size $[N_t N_{vs} N_t \times 1]$ and $r_t$ is a vector of size $[N_t N_{vs} N_t \times 1]$ obtained by concatenating traces for all available pairs of sources (or receivers) and virtual sources. Finally, $P^+_f$ is a linear operator applying MDC in the time domain as detailed in [35].

Moreover, to mitigate the ill-posed available pairs of sources (or receivers) and virtual sources. Finally, $P^+_f$ of size $[N_t \times N_t]$ is a matrix of size $[N_t \times N_t]$, and $R_f$ is a matrix of size $[N_t \times N_t]$. This system of equations can be solved with direct methods [9], [16] or iterative Block–Krylov solvers [26]. Alternatively, a time-domain formulation can be written as [20]

$$r_t = \arg\min_{r_t \in \mathcal{A}} \frac{1}{2} \| p_t^+ - P^+_f r_t \|_2^2$$

(3)

where $p_t^+$ is a vector of size $[N_t N_{vs} N_t \times 1]$ and $r_t$ is a vector of size $[N_t N_{vs} N_t \times 1]$ obtained by concatenating traces for all available pairs of sources (or receivers) and virtual sources. Finally, $P^+_f$ is a linear operator applying MDC in the time domain as detailed in [35].

Moreover, to mitigate the ill-posed nature of MDD, the solution can be forced to belong to a subspace of interest, $\mathcal{A}$, based on various physical constraints (e.g., reciprocity, causality) that the wavefield is expected to satisfy. As discussed in detail in [19] and [20], the constrained least-squares problem in (3) can be conveniently converted into an equivalent unconstrained least-squares problem

$$z = \arg\min_z \frac{1}{2} \| p_t^+ - \tilde{P}^+_f z \|_2^2$$

(4)

where $\tilde{P}^+_f = P^+_f A$. A linear preconditioner $A$ is used to ensure that the estimated reflectivity $r_t = Az$ is projected onto the set of admissible solutions in $\mathcal{A}$. In general, when multiple physical constraints are available, a chain of linear preconditioners can be used. For example, by choosing $A = A_0 A_c$, we can enforce both reciprocity (i.e., $R(x_R, x_V, t) = R(x_V, x_R, t)$) through $A_0$ and causality via $A_c$. We refer the reader to [19] for a detailed description of the implementation of such preconditioners. The unconstrained least-squares functional in (4) can be minimized by means of gradient-based iterative solvers (e.g., LSQR—[28]). For simplicity, in the remainder of the article, we will refer to the solution of (4) as full-gradient MDD.

However, since the forward problem in (1) holds for each of the available sources [9], in this work, we suggest to write the functional in (3) into an equivalent form. More precisely, the proposed functional is written in the finite-sum form

$$r_t = \arg\min_{r_t \in \mathcal{A}} \frac{1}{N_v} \sum_{k=1}^{N_v} f_{t,k}(r_t)$$

(5)

where each function $f_{t,k}$ is convex and L-Lipschitz continuous. In this specific case, we choose $f_{t,k}$ to be

$$f_{t,k}(r_t) = \frac{1}{2} \| p_{t,k}^- - P^+_{t,k} r_t \|_2^2$$

(6)

where $p_{t,k}^-$ is a vector of size $[N_{vs} N_t \times 1]$ that contains the up-going wavefield originated from the $k$th source. Similarly, $P^+_{t,k}$ is a linear operator performing MDC with the down-going wavefield of the $k$th source. Note that the constraint in (5) can be once again turned into a physical preconditioner similar to that in (4) by substituting $P^+_{t,k}$ with $\tilde{P}^+_{t,k} = P^+_{t,k} A$.

Drawing upon the vast literature of stochastic optimization [29], [30], the inverse problem in (5) can be conveniently solved using the stochastic gradient descent algorithm. Note that this hings on the fact that we have been able to recast the original problem 2 into an equivalent finite-sum form (5).

However, while vanilla stochastic gradient descent requires updating the model parameters $r_t$ at each iteration using an inexact gradient from a single, randomly selected function $f_{t,k}$, in this work, we rely on a more efficient mini-batch version. More specifically, rather than computing the exact gradient of (5) at each iteration (or the gradient from a single source) and using it to update the current $r_t$ vector, the available sources are grouped in batches of size $N_{sb} < N_t$ (see Fig. 1) and the gradient of each batch $\nabla_S$ is computed as follows:

$$\nabla_S = - \sum_{t_s \in S_{batch}} P^+_{t_s}(p_{t_s}^+ - P^+_f r_t)$$

(7)

or using $\tilde{P}^+_f$ for the preconditioned problem. Note that in practice, this gradient is more efficiently computed by creating a new operator and data vector as follows:

$$\begin{bmatrix} P^+_{t_1} & \cdots & P^+_{t_{Sbatch}} \\ \vdots & \ddots & \vdots \\ P^+_{t_{Sbatch}} & \cdots & P^+_{t_{Sbatch}} \end{bmatrix} \begin{bmatrix} p_{t_1}^- \cdots p_{t_{Sbatch}}^- \end{bmatrix}$$

(8)

and simply performing $\nabla_S = - P^+_{t_{Sbatch}}(p_{t_{Sbatch}}^- - P^+_f r_t)$. This approximated gradient is then used in the stochastic gradient descent (SGD) and the SGD with Nesterov Momentum (N-SGD) iterations as follows:

\begin{algorithm}
\caption{Stochastic Gradient Descent (SGD)}
\begin{algorithmic}
\State Initialize: $r_t, \alpha$
\For {epoch = 0 : $N_{epochs}$}
\For {batch = 0 : $N_{batch}$}
\State $r_t \leftarrow r_t - \alpha \nabla_S$
\EndFor
\EndFor
\end{algorithmic}
\end{algorithm}

\begin{algorithm}
\caption{SGD With Nesterov Momentum (N-SGD)}
\begin{algorithmic}
\State Initialize: $r_t, b_t, \alpha, \mu$
\For {epoch = 0 : $N_{epochs}$}
\For {batch = 0 : $N_{batch}$}
\If {epoch = 0}
\State $b_t \leftarrow \nabla_S$
\Else
\State $b_t \leftarrow \mu b_t + \nabla_S$
\EndIf
\State $r_t \leftarrow r_t - \alpha (\nabla_S + \mu b_t)$
\EndFor
\EndFor
\end{algorithmic}
\end{algorithm}

where the implementation of Nesterov momentum is based on [31]. In the following, the solution of (5) will be referred to as stochastic MDD. Moreover, the word epoch is used when referring to a set of iterations that utilize all of the available sources in a dataset, while iteration or step is used to indicate a single gradient step computed with a batch of sources. In this work, we select batchs without replacement; however, batches are not fixed throughout the inverse process, rather they are redrawn at every epoch (i.e., different sources will be part of different batches at different epochs). Based on this, when comparing the solutions of different algorithms, we will therefore assume that the cost associated with one epoch of a
mini-batch algorithm is about the same as that of one iteration of a full-gradient algorithm. Therefore, to avoid confusion, we also use the word epoch when referring to a single gradient step in full-gradient MDD. This observation is based on the fact that by considering a subset of sources at each step of the stochastic solvers, the computational cost of applying the operator $P_{S_{\text{batch}}}$ and its adjoint is also reduced by roughly the ratio of the overall number of sources over that of the sources belonging to a batch (see Appendix A for a more detailed analysis of the computational cost of the two MDD algorithms).

III. Examples

Four numerical examples are discussed in this section as depicted in Fig. 2. The first and last two examples are similar in that they use seismic data after up/down decomposition as input to the MDD problem. In these cases, we aim to create a virtual dataset that is free from surface-related multiples. While the decomposition process is likely to introduce some leakage of the up-going component into the down-going and vice versa, we expect such errors to be minimal. Moreover, in all of these examples, the overburden is very simple (i.e., free-surface); while this creates reverberations in the up- and down-going wavefields that translate into multiple events at the nonzero time-space lag in the point spread function of the MDD operator (i.e., $P_{f}^{+}H_{f}P_{f}^{-}$), it does not introduce complex illumination imbalance to be corrected by the deconvolution process. Finally, a wide-aperture receiver array is available.

In the second synthetic example instead, MDD is performed on up- and down-going wavefields that have been redatumed to a certain depth level by means of a data-driven redatuming process, namely SRM redatuming. This process inevitably introduces coherent artifacts in both the up- and down-going wavefields, as extensively discussed in [20]. Moreover, the overburden to be compensated for by MDD is highly complex, including a high impedance salt body. Finally, the aperture of the receiver array is limited, meaning that a number of contributions present in the upgoing wavefield may not be reconstructed by the MDD modeling operator. The associated MDD problem is, therefore, more unstable than in the previous examples and requires physical preconditioning to produce satisfactory results.

In all of the experiments, qualitative and quantitative assessment of the results produced by the full-gradient and stochastic MDD algorithms will be presented. As far as the latter is concerned, when dealing with synthetic data, the root mean-square error (RMSE) is used as metric

$$\text{RMSE}(r_{\text{true}}^{r}, r_{\text{est}}^{r}) = \sqrt{\frac{\|r_{\text{true}}^{r} - r_{\text{est}}^{r}\|_{2}^{2}}{\max \{|x_{1}|, |x_{2}|, \ldots, |x_{N}|\}}}$$

where $r_{\text{true}}^{r}$ is the true reflection response obtained by means of finite-difference modeling, and $r_{\text{est}}^{r}$ is the estimated reflection response by means of MDD. Here, $\|x\|_{2}^{2} = (\sum_{i=1}^{N} x_{i}^{2})^{1/2}$ is the squared Euclidean distance, and $\max(|x|) = \max(|x_{1}|, |x_{2}|, \ldots, |x_{N}|)$.

A. Synthetic OBC Up/Down Deconvolution

The first numerical example deals with the variable-density, variable-velocity model shown in Fig. 2(a). An OBC acquisition setup with receivers placed along a dipping seabed.
(1.9° dip) is used to create the dataset. This example is partially inspired by the work of [13] that showed the importance of deconvolving the up- and down-going wavefields in a multidimensional fashion in the presence of a nonflat seafloor. Note that to ensure minimal error in the wavefield separation step, both the velocity and density of the seabed are chosen to be close to those of the water layer (i.e., soft seafloor). However, a hard dipping layer is introduced at around 600 m depth; this reflector represents the main generator of strong free-surface multiples that we wish to tackle by means of MDD. A line of 201 sources at depth $z_s = 10$ m with spatial sampling $dx_s = 15$ m is used alongside a line of 201 receivers at varying depth ranging from $z_r, \min = 328$ m to $z_r, \max = 428$ m and spatial sampling $dx_r = 15$ m. A Ricker wavelet with central frequency $f_c = 20$ Hz is used to model the pressure and vertical particle velocity data and a free surface is added on top of the model. The dual-sensor data are subsequently separated into their up- and down-going pressure components in the $(\omega - k_r)$ domain [32], as shown in Fig. 3(a) and (b). The ideal reflection response from sources and receivers at the seafloor and in the absence of both the free surface and water column is also modeled and displayed in Fig. 3(c).

In the first experiment, we perform redatuming for a single virtual source in the middle of the receiver array ($x_{sv} = 1500$ m). Given the high quality of the input wavefields and overall low level of noise in the data, additional constraints that couple the responses for multiple virtual sources are not required; apart from an improvement in terms of computational speed [35], similar results are in fact expected when inverting for multiple virtual sources simultaneously. The redatumed response obtained by applying the adjoint of the modeling operator to the up-going data in Fig. 3(a) is displayed in Fig. 4(a). Since the adjoint of the modeling operator can be equivalently interpreted as the cross correlation between the up- and down-going wavefields, the resulting wavefield contains a mix of physical events with correct kinematic and spurious events due to the crosstalk between unrelated events in the up- and down-going wavefields [9]. More specifically, when compared to the ideal reflection response in Fig. 3(c), this wavefield contains strong spurious events at around 1 s and 1.75 s that we wish to remove when solving the inverse problem in (3) or (5).

Three different inversions are carried out and shown in Fig. 4(b)–(d). Fig. 4(b) represents the benchmark reflection response obtained by means of full-gradient MDD after 10 iterations of LSQR. The following two panels display the reflection responses estimated by means of N-SGD for 10 and 20 epochs, respectively. In all cases, a batch size of $N_b = 32$ is used. Note that since we have access to 201 sources, the dataset is divided into $\lfloor N_s / N_b \rfloor + 1 = 7$ batches, where the last batch contains only nine sources. Alternatively, we could have aggregated these nine sources to the penultimate batch. Finally, the step-size selection is based on the semi-heuristic criterion described in Appendix B. Visually, the reflection responses retrieved by the various algorithms show a good agreement with the ideal reflection response in Fig. 3(c). Importantly, the strong spurious events in Fig. 4(a) have been successfully attenuated, while the underlying weaker primary events are now visible. A closer inspection of the results, alongside a quantitative assessment of the reconstruction error of the various algorithms [see Fig. 4(f)], reveals a faster convergence of the stochastic MDD algorithms over the full-gradient MDD method. More specifically, if we compare the error norms after ten epochs, we observe that both SGD and N-SGD produce a solution that is of superior quality to that of full-gradient MDD. Although the quality of the reflection response for full-gradient MDD further improves with the number of epochs and overtakes that of SGD at around 20 iterations, the result of N-SGD is still the one that produces the lowest RMSE. Such favorable behavior can be explained by the fact that for each epoch of N-SGD we are performing seven inexact gradient steps versus a single, although optimal, gradient step for LSRQ in full-gradient MDD. Moreover, to assess the possible impact that the choice of different batches may have on the final solution, we run 30 independent inversions with the N-SGD solver using different initial random seeds. The shaded blue area in Fig. 4(f) shows that the proposed algorithm is pretty much insensitive to the seed parameter. Finally, for this specific example, similar behavior is also observed for the norm of the residual, where the stochastic algorithms show a faster convergence to zero residual. Note that both iteration-wise norms (thin lines) and epoch-averaged norms (thick lines) are displayed in Fig. 4(e). Nevertheless, as we will see in later examples, the norm of the residual does not represent a reliable proxy metric for the error norm, which in real-life examples cannot be directly evaluated.

Finally, we want to assess the impact of batch size on the convergence of the stochastic MDD algorithms. This parameter plays a key role in that a too small batch size may produce highly corrupted gradients as well as introduce computational overhead due to the amount of forward and inverse fast-Fourier transforms (FFT’s) additionally required in the evaluation of the gradients for the stochastic algorithms (see Appendix A for details). Nevertheless, a too large batch size may use redundant information in the generation of the gradients; moreover, although not directly shown in this article, in 3-D applications, this may lead to out-of-core operations as
Reflection response estimates. (a) Cross correlation (i.e., adjoint), (b) full-gradient MDD after 20 epochs, (c) stochastic MDD with N-SGD after ten epochs, and (d) stochastic MDD with N-SGD after 20 epochs. (e) Residual and (f) error norms as a function of epochs for the full-gradient MDD (black lines), MDD by means of SGD (red lines), and MDD by means of N-SGD (blue lines). In both plots, thick solid lines with dots display the values of the respective norms at the end of each epoch. Thin solid lines in panel (e) represent the norms at each iteration for the stochastic MDD algorithms (i.e., contain $N_s/N_{\text{batch}}$ elements per epoch). A blue shaded region in panel (f) represents two standard deviations across 30 choices of the initial random seed. Note that the norms of the stochastic MDD algorithms start from the first step and intermediate values for the iterations up to the end of the first epoch are also displayed, while those of the full-gradient MDD start from $n_{\text{epochs}}=1$.

Gradients for (a) full source array, (b) batch of 64 randomly selected sources, and (c) batch of 32 randomly selected sources. (d) Error norms as function of epochs for the various algorithms.

Down- and up-going common-receiver gathers for a receiver in the middle of the subsurface array. (c) Ideal reflection response for a virtual source in the middle of the subsurface array.

for the full source array, and batches of 64 and 32 randomly selected sources. It is evident that the main features of the full gradient are preserved in the approximate gradients; nevertheless, some spurious hyperbolic events with opposite curvature arise due to the poor sampling of stationary points. The behavior of the error norms (see Fig. 5) does, however, indicate that the smaller number of sources used to compute each gradient, the faster the overall convergence, leading to a tradeoff between computational efficiency in the modeling operator and convergence speed in the choice of the batch size.

B. Synthetic Subsalt Redatuming

In this second example, we consider a scenario of source-side, target-oriented redatuming below a salt body [see Fig. 2(b)]. The acquisition geometry comprises of a line of 201 sources at depth $z_s = 50$ m with spatial sampling $dx_s = 40$ m and 151 receivers at depth $z_r = 4500$ m with spatial sampling $dx_r = 20$ m. Here, the input up- and down-going wavefields are not simply obtained by means of wavefield separation, rather they are the product of a step of
receiver-side redatuming of surface data. More specifically, the up- and down going separated surface data are used as input to the SRM redatuming scheme [21]; a highly ill-posed inverse problem that despite its high degree of accuracy when compared to other redatuming methods (e.g., [4], [33]) inevitably introduces errors in the form of both coherent and incoherent noise in the redatumed responses. Fig. 6 displays the down- and up-going wavefields for a receiver in the middle of the subsurface array, alongside the ideal reflection response deprived of any overburden effect (i.e., modeled in a truncated medium starting from a depth of $z = 4500$ m).

As shown in [20] (their Fig. 3), the up-going wavefield used as data to be matched by the MDD equations greatly differs from its “modeled” counterpart, obtained by applying (1) to the redatumed down-going field and the true reflection response. Moreover, because of the presence of a small-aperture receiver array, differences inevitably arise between the modeled wavefield and the field to match. Because of these reasons, this dataset presents a perfect playground for the development of robust MDD algorithms; we now aim to assess the capabilities of stochastic MDD in the presence of noise in both the data and modeling operator, when inverting for either one or multiple virtual sources. The former approach presents the clear advantage of being less computationally demanding and easier to parallelize over multiple virtual sources; the latter provides the opportunity to introduce a reciprocity physical constraint, however, it comes with a number of additional computational challenges due to the size of the model and data vectors to be used in the inversion process. The redatumed wavefields obtained using a single virtual source in the middle of the array are displayed in Fig. 7. The response retrieved by means of full-gradient MDD [see Fig. 7(a)] is clearly contaminated by strong dipping noise, which arises due to the fact that both the data and modeling operator are inexact. By computing the norm of the model error through iterations, we can see how the black line in Fig. 7(e) decreases in the
the previous result obtained by using a single virtual source. The reflection responses from the stochastic MDD methods are also very clean and of slightly higher quality when compared to their full-gradient counterpart. Fig. 9 shows a close-up of the error norms for the six different inversion results shown in Figs. 7 and 8. It can be observed how, in general, the introduction of a reciprocity preconditioner improves the solution of both full-gradient and stochastic MDD. However, while this seems to be of vital importance for the former method, the improvement is fairly minor in the latter, 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 great importance for the application of MDD to large-scale, 3-D datasets that were solved for the entire set of virtual sources at the same time may be beyond the reach of our current compute capabilities [34], [35].

C. Volve OBC Up/Down Deconvolution

In this final example, we consider a 2-D line of the open-source Volve dataset. Volve is an oil field located in the central part of the North Sea, five kilometers north of the Sleipner Øst field. The field was shut down in 2016, with the facility removed in 2018, and the historical subsurface and production data were made available by Equinor and partners in June 2018.

In order to be able to assess the performance of stochastic MDD against full-gradient MDD, we begin by creating a synthetic, Volve-like dataset. The dataset creation is composed of three main steps.

1) The migration velocity model in Fig. 2(d) \((ST10010ZC11-MIG-VEL_MIG_VEL_3D.JS-017527.segy)\) is first converted into an equivalent acoustic impedance background model with the help of available well logs by means of linear regression.

2) The 3-D post-stack seismic dataset \(ST10010ZC11_PZ_FSDM_KIRCH_FULL_D.MIG_FIN.POST_STACK.3D.JS-017536.segy\) is used alongside the background acoustic impedance model as input to a step of post-stack inversion (see [35] for more details); the resulting detailed acoustic impedance model is converted back to an equivalent velocity model using the inverse of the velocity-acoustic impedance relation at well locations.

3) Pressure and particle velocity seismic data are modeled using a Ricker wavelet with central frequency \(f_c = 20\) Hz and a free surface at the top of the model. A line of 110 sources at depth of \(z_s = 6\) m with spatial sampling \(dx_s = 50\) m is used alongside a line of 180 receivers at varying depth ranging from \(z_{r,min} = 86\) m to \(z_{r,max} = 99\) m and spatial sampling \(dx_r = 25\) m. Similarly, the reference reflection response is modeled by filling the water layer with the velocity of the seafloor and deactivating the free surface. In this case, both sources and receivers are placed along the seafloor [see Fig. 10(c)].

Similar to previous experiments, the up- and down-going pressure wavefields are created by means of wavefield
Fig. 11. Reflection response estimates after 30 epochs. (a) Cross correlation. (b) Single-virtual-source full-gradient MDD. (c) Multi-virtual-source full-gradient MDD. (d) Single-virtual-source stochastic MDD with N-SGD. (e) Multi-virtual-source stochastic MDD with N-SGD.

Fig. 12. Close-up of error norms for the different algorithms. Solid and dashed lines are used for single- and multi-virtual strategies, respectively.

separation [see Fig. 10(a) and (b)] and used as input to the various MDD algorithms. It must be noted here that, while wavefields are of much greater complexity than those in the first example, the amount of noise introduced by the wavefield separation process is negligible when compared to that introduced by a previous step of redatuming, like in the subsalt example. In other words, we expect both full-gradient and stochastic MDD to perform well in this scenario; our main interest lies in understanding the stability and convergence of the various MDD processes as a proxy to what we will observe in the field data experiments (without a reference solution to compare our results onto). Fig. 11 displays the reconstructed reflection responses for a virtual source in the middle of the receiver array using the adjoint of the modeling operator [see Fig. 11(a)], single- and multi-virtual source full-gradient MDD [see Fig. 11(b) and (c)], and single- and multi-virtual source stochastic MDD [see Fig. 11(d) and (f)]. All inversions are carried out with a causality preconditioner (i.e., a space–time mask that mutes amplitudes above the direct arrival in the reconstructed reflection responses), and an additional reciprocity preconditioner is used for multi-virtual source inversions. Moreover, for the stochastic gradient methods, the dataset is divided into four batches of $N_{sb} = 32$ sources. Given the presence of a fairly shallow seabed, crosstalk between unrelated primaries and free-surface multiples is visible in the reflection response constructed by means of cross correlation. Such spurious events are clearly suppressed in the MDD results independent of the choice of the algorithm. Nevertheless, given the complexity of the wavefield and therefore the ill-posed nature of the problem, significant incoherent noise is introduced in the solution obtained from single-virtual-source full-gradient MDD [see Fig. 11(b)]. Such noise is visibly reduced when introducing a reciprocity preconditioner as part of the multi-virtual-source full-gradient MDD [see Fig. 11(c)]. The reflection responses obtained by means of stochastic MDD corroborate our previous findings that working with batches of sources seems to act as a natural regularizer to the final solution. Once again, working with multiple virtual sources and introducing a reciprocity preconditioner further improves the quality of the solution; however, the improvement in this case is much less evident. Similar observations can be drawn from Fig. 12, where we observe the usual semi-convergence behavior for the full-gradient MDD solutions and a more stable convergence for the stochastic MDD cases. Nevertheless, in this specific example, the convergence of full-gradient MDD, especially when using multiple virtual sources in combination with a reciprocity preconditioner, seems to slightly outperform that of the stochastic MDD algorithms.

Moving onto the field dataset, a 2-D line of sources and receivers is selected from the ST10010_1150780_40203.sgy file that contains a portion of the 3-D OBC dataset acquired by Statoil in 2010 [see Fig. 13(a)]. The pressure and vertical particle velocity components are preprocessed using a similar flow to that described in [10] and [36], and the resulting up- and down-going wavefields for a source in the middle of the
Fig. 13. (a) Acquisition geometry with selected 2-D line of sources (red dots) and receivers (blue dots). (b) Down-going and (c) up-going pressure data for a single source in the middle of the 2-D line (large red dot in panel a).

2-D acquisition geometry are displayed in Fig. 13(b) and (c). Full-gradient and stochastic MDD is applied to the separated wavefield for 40 epochs. Similar to the synthetic example, causality and reciprocity preconditioners are used to aid the solution of both MDD algorithms. Fig. 14 displays the estimated reflection responses by means of cross correlation [see Fig. 14(a)] and various MDD algorithm [see Fig. 14(b)–(e)]. Note that while the different MDD algorithms are all able to some extent to remove the crosstalk artifacts visible in the cross correlation gather, the stochastic MDD algorithms produce cleaner wavefields (see, e.g., the event indicated by red arrows). The quality of the different reconstructions can also be appreciated in a close-up taken from around 0.8 s (see Fig. 15). Finally, to better assess 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 all imaged by means of reverse-time migration (see Fig. 16). 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.
Fig. 15. Close-up of the reflection responses. (a) Cross correlation. (b) Single-virtual-source full-gradient MDD. (c) Multi-virtual-source full-gradient MDD. (d) Single-virtual-source N-SGD MDD. (e) Multi-virtual-source N-SGD MDD.

Fig. 16. Images obtained using (a) full pressure data, (b) up-going pressure wavefield, (c) reflection response from multi-virtual-source full-gradient MDD, and (d) reflection response from multi-virtual-source N-SGD MDD.

Fig. 17. (a) Eigenvalue spectrum (image) and average frequency spectrum (solid line) for the synthetic OBC down-going wavefield in the frequency domain. A vertical dashed line indicates the frequency associated with the peak of the amplitude spectrum (and the maximum eigenvalue). (b) Residual norm and (c) step size as a function of iterations for the steepest descent algorithm (solid), the Landweber algorithm with step size smaller (dashed), and larger (dashed and dotted) than the maximum allowed value.

IV. DISCUSSION

Stochastic optimization algorithms have been popularized in geophysics by [37] and [38] in the context of partial differential equation-based (PDE-based) nonlinear inverse problems, that is, full waveform inversion (FWI). In such a scenario, the cost of computing a gradient is driven by the number of PDE solves required to evaluate a chosen objective function; the benefit of using inexact gradients is represented by the fact that nearby sources provide highly redundant information in the construction of the gradient itself. Moreover, as the problem is inherently nonconvex, there is no guarantee to reach the global minimum irrespective of the fact that full or partial gradients are used.

In our work, we are instead dealing with a linear inverse problem, which ideally has a unique minimizer; this may cast some doubts about the benefit of using stochastic solvers. It is, however, important to remember that the forward operator of the MDD problem is highly ill-posed—that is, the kernel matrix is rank deficient and has a large null space as shown, for example, in Fig. 17(a). Therefore, a variety of solutions exist that match the data equally well. Moreover, recent research in the medical imaging community has also shown that a linear...
inverse problem like magnetic resonance imaging can greatly benefit from the use of stochastic gradient algorithms [39], [40]. Additionally, similar to FWI, the information carried by each source to the gradient of the MDD problem can be redundant. These two factors combined motivate the use of stochastic optimization within the context of time-domain MDD. Our numerical experiments highlight two main benefits of stochastic gradient over full gradient optimization for MDD: robustness to noise in the data and operator, and faster convergence at times. The latter benefit is clearly highlighted in the first synthetic example where high-fidelity up- and down-going wavefields are used as input to the MDD problem. As negligible noise affects such measurements (even after wavefield separation), both the full and stochastic solvers converge to an accurate solution; nevertheless, the convergence of the N-SGD solver is faster by a factor of nearly two in comparison to full-gradient MDD. The former is more evident in the second example where noisy measurements (from SRM redatuming) are used to form both the data term and modeling operator. The use of stochastic gradients acts as a natural regularizer reducing the reliance on expensive preconditioners and/or stopping criteria.

A major difference between full gradient and stochastic gradient-based algorithms is that the former optimizes a constant functional, while the cost function changes from step to step in the latter. Although the use of large batch sizes can provide an unbiased estimate of the gradient, no analytical, exact formula exists for the step size at each iteration. This is obviously not the case in full-gradient iterative solvers, like CGLS [41] or LSQR. The choice of the step size becomes therefore a very important hyperparameter to the success of stochastic MDD. The problem here is twofold: first, we wish to find the best (or a satisfactory) step size that provides fast and, at the same time, stable convergence. Second, we would like to be able to identify this scaling factor upfront without the need for trial-and-error that would inevitably narrow the computational gap between the full and stochastic MDD. The semi-heuristic procedure based on the Landweber iteration (see Appendix B) is shown in the numerical examples to be effective in this context.

V. CONCLUSION

We have presented a novel formulation of MDD that recast the associated inverse problem into the optimization of a finite-sum functional. By doing so, its solution is obtained by means of stochastic gradient descent algorithms, allowing gradients at each step to be 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. As a by-product of an overall more stable convergence, stochastic MDD is also shown to rely less heavily on inter-virtual-source constraints (i.e., reciprocity preconditioning), ultimately easing its extension to large-scale, 3-D datasets.

APPENDIX A

COMPUTATIONAL COST OF STOCHASTIC MDD

In this Appendix, we compare the computational cost of full gradient and stochastic MDD. It must be noted that, unlike FWI based on random shot selection, our modeling operator is only partially separable among shots and therefore some additional overhead arises when splitting it into a group of sources.

To being with, let us revisit the computational cost of the MDC operator in (1). This operator is composed of three consecutive steps, namely forward Fourier transform (along the time axis), batched matrix–matrix multiplication (MMM)—which reduces to matrix–vector multiplication (MVM) for the single-virtual-source case—and inverse Fourier transform (along the time axis), whose computational costs are

\[
\text{FFT: } \mathcal{O}(N_t \log_2(N_t)/2 \cdot N_r \cdot N_{vs})
\]

\[
\text{Batched MMM: } \mathcal{O}((N_t/2) \cdot N_r \cdot N_v \cdot N_{vs})
\]

\[
\text{IFFT: } \mathcal{O}((N_r \log_2(N_r)/2) \cdot N_t \cdot N_{vs})
\]

where we assume here for simplicity that the number of samples in the Fourier domain equals half of those of the time-domain series. This stems from the fact that seismic data are real-valued functions and therefore a real-FFT algorithm [42] can be used.

In the case of full-gradient MDD, the cost of a single epoch is therefore equal to twice the cost of the MDC operator, as a forward and adjoint pass is required at each step of the solver.

\[
\text{Full MDD (1epoch): } 2 \cdot \left[ \mathcal{O}((N_t \log_2(N_t)/2) \cdot N_r \cdot N_{vs}) + \mathcal{O}((N_t/2) \cdot N_r \cdot N_v \cdot N_{vs}) + \mathcal{O}((N_r \log_2(N_r)/2) \cdot N_t \cdot N_{vs}) \right].
\]

On the other hand, the cost of an epoch of stochastic MDD equals \(2N_b\) the cost of an MDC operation with \(N_{sb}\) sources, where \(N_b\) is the overall number of batches of sources. In this case, \(N_{sb}\) is also used in spite of \(N_t\) in the definition of the cost of the Batched MMM and IFFT steps of MDC operator. The cost of the FFT step is, however, the same as for the full MDC operator as it is independent on the source axis. An explicit expression for the cost of an epoch of stochastic
MDD is

Stochastic MDD (1 epoch):

\[ 2 \cdot N_b \cdot \left[ O((N_A \log_2(N_b))/2) \cdot N_r \cdot N_v \right] \]
\[ + O((N_A/2) \cdot N_{sb} \cdot N_r \cdot N_v) \]
\[ + O((N_A \log_2(N_b))/2) \cdot N_{sb} \cdot N_v) \]
\[ = 2(N_b - 1) \cdot O((N_A \log_2(N_b))/2) \cdot N_r \cdot N_v \]
\[ + \text{Full MDD (1 epoch)}. \]  

(12)

This result shows that an additional \( 2(N_b - 1) \) forward Fourier transforms are performed at each epoch of the stochastic MDD algorithm compared to its full-gradient counterpart. However, when \( N_A \gg \log_2(N_b) \), the second step of the MDC operator dominates in terms of computational cost. This is a common scenario as the number of time samples of a seismic recording is usually in the order of \( 10^3 \) (so that \( \log_2(N_b) \sim 10 \)), while the number of sources can be around \( 10^2 - 10^3 \) for 2-D acquisition systems, and \( 10^3 - 10^5 \) for 3-D acquisition systems.

**APPENDIX B**

**SELECTING STEP SIZE IN STOCHASTIC GRADIENT DESCENT**

A key factor in the success of stochastic MDD is represented by the choice of the step size. A too small step size would inevitably lead to slow convergence, while a divergent solution would be produced when choosing a too large step size. Obviously, this parameter could be optimized via a trial-and-error strategy; however, such a choice would require running a couple of iterations while monitoring the residual norm until a successful step size is found. These additional gradient computations would impact the overall cost of the inversion.

Alternatively, we propose to use a semi-heuristic procedure based on the well-known Landweber iteration [43], a variant of the gradient descent algorithm with fixed step size. Simply put, given a linear inverse problem (e.g., \( x = \arg \min_y \{ \frac{1}{2} ||y - Ax||^2 \} \)), Landweber showed that convergence is guaranteed for the following choice of the step size, \( 0 < \alpha < 2/\sigma^2(A) \).

Here, \( \sigma(A) \) is largest singular value of the matrix (or operator) \( A \). Such a condition can be easily verified with a numerical example

\[
A = \begin{bmatrix} 1 & 4 \\ 6 & 2 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

(13)

where \( \sigma^2(A) = 53 \). Fig. 18 shows the convergence behavior of the Landweber algorithm for two choices of step size: \( \alpha = 2.05\sigma^2(A) \) (blue line) and \( \alpha = 1.9\sigma^2(A) \) (red line). It is clear that convergence to the solution cannot be achieved when the above condition is not met, even when \( \alpha \) is chosen to be only 2.5% larger than the maximum allowed value. Finally, the Landweber iteration is also compared to the steepest descent algorithm (green line), where the step size varies at each iteration and is chosen to minimize the cost function along the selected gradient direction. While this selection criterion leads to a much faster convergence, it cannot be easily extended to stochastic gradient descent algorithms.

Moving back to the MDD problem, a question arises about how to compute the largest singular value of the MDC operator. While estimating it by means of matrix-free power iteration methods is very expensive as it requires the evaluation of several forward and adjoint passes of the operator, [19] showed that the singular values of the operator are equivalent to those of the kernel (i.e., frequency-domain down-going wavefield). Furthermore, as shown in Fig. 17(a), the largest singular value of the operator is associated with the frequency matrix with the strongest contribution to the data; this can be easily estimated by looking at the average spectrum of down-going wavefield [black curve at the bottom of Fig. 17(a)]. Fig. 17(b) shows the residual norm as a function of iterations when solving (3) with the Landweber iteration using a step size equal to \( 1.9/\sigma(P^{-H}P) \) and \( 2.05/\sigma(P^{-H}P) \) as well as the analytical optimal step size.

Heuristically, we found that for stochastic MDD, the choice of the step size must be made more conservatively due to the fact that inexact gradients are used at each step; in other words, by using a step size of 40%–60% of the theoretical upper limit for the Landweber iteration, we are able to consistently produce stable solutions across various numerical examples presented in this article.

**ACKNOWLEDGMENT**

All numerical examples can be accessed at https://github.com/DIG-Kaust/MDD-StochasticSolvers. Moreover, the synthetic and real Volve datasets can be accessed at https://github.com/DIG-Kaust/VolveSynthetic and https://data.equinor.com, respectively. The authors are grateful to Equinor and partners for releasing the Volve dataset. Matteo Ravasi would like to thank Claire Birnie (KAUST) for their insightful discussions.

**REFERENCES**

[1] E. Biondi, B. Biondi, and G. Barnier, “Target-oriented elastic full-waveform inversion through extended-migration redatuming,” in Proc. 89th Annu. Int. Meet. SEG Tech. Program. Expanded Abstr., pp. 1228–1232, Aug. 2018, doi: 10.1190/segam2018-2998407.1.

[2] Q. Guo and T. Alkhalifah, “Target-oriented waveform redatuming and high-resolution inversion: Role of the overburden,” Geophys. J., vol. 85, no. 6, pp. R525–R536, Nov. 2020.

[3] Y. Li and T. Alkhalifah, “Target-oriented high-resolution elastic full waveform inversion using redatumed multi-component data,” in Proc. 82nd Annu. Conf. Expo., EAGE, Expanded Abstr., 2021, pp. 1–5, doi: 10.3997/2214-4609.202112858.

[4] K. Wapenaar, J. Thorbecke, J. van der Neut, F. Broggini, E. Slob, and R. Snieder, “Marchenko imaging,” Geophys. J., vol. 79, no. 3, pp. WA39–WA57, 2014, doi: 10.1002/ggj.203.

[5] J. van der Neut, I. Vasconcelos, and K. Wapenaar, “On Green’s function retrieval by iterative substitution of the coupled Marchenko equations,” Geophys. J. Int., vol. 203, no. 2, pp. 792–813, Nov. 2015, doi: 10.1093/gji/ggv330.

[6] M. Ravasi, “Rayleigh-marchenko redatuming for target-oriented, true-amplitude imaging,” Geophys. J., vol. 82, no. 6, pp. S439–S452, Nov. 2017, doi: 10.1093/gji/ggw355.

[7] X. Li, T. Becker, M. Ravasi, J. Robertsson, and D.-J. van Manen, “Closed-aperture unbounded acoustics experimentation using multidimensional deconvolution,” J. Acoust. Soc. Amer., vol. 149, no. 3, pp. 1813–1828, Mar. 2021, doi: 10.1121/10.0003706.

[8] L. Amundsen, “Elimination of free-surface related multiples without need of a source wavelet,” Geophys. J., vol. 66, pp. 327–341, 2001, doi: 10.1046/j.1365-246x.2001.05007.x.

[9] K. Wapenaar et al., “Seismic interferometry by crosscorrelation and by multidimensional deconvolution: A systematic comparison,” Geophys. J. Int., vol. 185, no. 3, pp. 1355–1364, Jun. 2011, doi: 10.1111/j.1365-246x.2011.05007.x.
