Pre-Stack Least-Squares Reverse Time Migration with an Exact Adjoint Operator for Ground-Penetrating Radar

Linan Xu, James Irving, Klaus Holliger

***Abstract*—This study develops pre-stack least-squares reverse time migration (LSRTM) for high-resolution ground-penetrating radar (GPR) imaging and compares its performance to standard RTM. We employ the Born approximation to derive the LSRTM forward operator and we utilize matrix operations to determine its exact adjoint. Tests on synthetic data confirm that LSRTM outperforms both pre-stack and post-stack RTM in terms of resolution and thus significantly improves the visibility of fine-scale structures. Further, the exact adjoint operator is shown to facilitate faster convergence and lead to a lower data misfit. Although LSRTM entails higher computational costs than standard RTM, we demonstrate that accelerated computing with GPU devices can make it remarkably affordable for 2D GPR imaging, thereby opening avenues for future applications in 3D.**

***Index Terms*—reverse time migration, discrete adjoint state method, Born approximation, least-squares migration, ground-penetrating radar, adjoint simulation.**

I. Introduction

GROUND-penetrating radar (GPR) is a prominent geophysical technique that uses high-frequency electromagnetic (EM) waves to image the shallow subsurface [1], [2]. Applications of GPR prevail in a wide variety of domains, including civil engineering (e.g., [3]), hydrogeology (e.g., [4]), glaciology (e.g., [5], [6]), archaeology (e.g., [7]), and sedimentology (e.g., [8]). Most commonly, GPR data are acquired in the form of surface-based common-offset reflection surveys, where a single transmitter and receiver antenna, spaced by some fixed distance, are moved along a profile line. One critical step in the processing of such data is migration, where the overall aim is to focus and relocate scattered and reflected wavefield components, recorded at the surface, back to the subsurface locations from which the energy was scattered. This, in turn, provides a map of subsurface reflecting interfaces corresponding to vertical contrasts in electromagnetic impedance. In this regard, the fidelity of the migration process is primarily determined by the accuracy of the associated wave propagation modelling. Unfortunately, the wave kinematics predicted by the most commonly employed migration methods for common-offset GPR data, notably ray-based Kirchhoff migration [9] and frequency-wavenumber-based Stolt migration [10], are accurate for media with smooth spatial velocity changes. Given that the shallow subsurface environment is often strongly heterogeneous, the use of such methods can result in significant errors. [11], [12].

For complex subsurface environments, reverse time migration (RTM) allows for a precise mapping of reflectors. RTM stands as the state-of-the-art imaging algorithm, employing numerical methods such as finite-differences, to model the EM wave propagation (e.g. [13], [14]). This does, however, come at significantly increased computational cost compared to conventional approaches. In this regard, post-stack RTM stands out as the most time-efficient implementation of RTM because it needs to solve the wave propogation problem only once. The method operates by propagating all-at-once the recorded data backwards in time at half the velocity of the medium, and extracting the reconstructed wavefield at time zero as the migrated image. The underlying physics are conceptualized by the exploding reflector [15] or superimposed point-diffractor [16] models. Post-stack RTM has been applied to GPR imaging since the 1990s [17], and has evolved through the adoption of more advanced forward modelling methods and algorithms (e.g., [18]–[21]). In principle, it is ideally suited to zero-offset data. However, the fixed transmitter-receiver antenna spacing for most GPR datasets means the zero-offset condition is not precisely satisfied. Further, post-stack migration does not allow for the preservation of true amplitudes [22], nor does it allow for analyses to be carried out with regard to resolution or the effects of inaccuracies in the migration velocity model. [23], [24].

Pre-stack RTM is a more general migration approach that is suitable for common- and multi-offset reflection data and can address many of the aforementioned issues. The corresponding algorithm consists of three key components: forward time propagation from the source, backward time propagation from the receiver, and application of an imaging condition. The imaging condition is pivotal in pre-stack RTM and has a strong influence on both the quality and robustness of the results obtained. The commonly used zero-lag cross-correlation imaging condition ([25], [26]) has proven to be particularly effective with multi-offset GPR data (e.g., [27]). Other more advanced imaging conditions have been developed within the exploration seismic industry (e.g., [28], [29]), and recent research has adapted them for GPR (e.g., [30], [31]). Despite these merits and advances, there are still challenges associated with pre-stack RTM, such as the presence of low-frequency artifacts, side-lobe smearing, and image-space aliasing caused by incomplete data (e.g., [32]–[34]). Directly attempting to suppress such artifacts after migration is generally impossible,

Dr. Linan Xu, Prof. James Irving and Prof. Klaus Holliger are with the Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland. E-mail: linan.xu.geophysics@gmail.com

Manuscript received December 17, 2023; revised Februry 16, 2024.

as differentiating between the artifacts and true reflection events in the image domain is inherently difficult.

In contrast to conventional migration, which essentially applies a linear operator to the data [35], least-squares migration (LSM) formulates reflection imaging as a linear inverse problem, where the aim is to find a migrated image that allows for satisfactory prediction of the observed data in a least-squares sense (e.g., [36]–[39]). By minimizing the data misfit, artifacts are reduced and the overall image quality is improved. Additionally, posing migration as an inverse problem allows for integrating a priori information and signal processing techniques into the procedure. The latter aids in achieving images with balanced amplitudes, reduced acquisition footprints, and enhanced resolution, while at the same time offering robustness against noise and inaccuracies in the migration velocity model (e.g., [40]–[47]). Least-squares reverse time migration (LSRTM) is a particularly prominent and powerful form of LSM. It has demonstrated a strong capacity for imaging complex geological scenarios in exploration seismology (e.g., [33], [45], [48]), but so far has not been explored in GPR research. Given the kinematic analogies that exist between seismic and GPR wave propagation [49], it is reasonable to assume that the adaptation and application of LSRTM to GPR data may prove to be as benenficial as it has been for seismic reflection data.

In this paper, we develop and explore pre-stack LSRTM for GPR imaging. First, we introduce the governing two-dimensional (2D) Maxwell equations, which is followed by a derivation of the forward operator using Born linearization. Next, utilizing the matrix-based adjoint-state method [50], we derive the exact adjoint of the forward operator, with a specific focus on the convolutional perfectly matched layer (CPML) boundary conditions [51]. Finally, we proceed to evaluate the efficacy of LSRTM in the context of GPR imaging.

II. Governing equations for 2D GPR simulation

Following [14], we begin with Maxwell's curl equations in the frequency domain

$$\nabla \times \vec{\tilde{\boldsymbol{E}}} = -i\omega\mu\vec{\tilde{\boldsymbol{H}}}, \tag{1}$$

$$\nabla \times \vec{\tilde{\boldsymbol{H}}} = (\sigma + i\omega\epsilon)\vec{\tilde{\boldsymbol{E}}}, \tag{2}$$

where $\vec{\tilde{\boldsymbol{E}}} = \tilde{E}_x, \tilde{E}_y, \tilde{E}_z$ and $\vec{\tilde{\boldsymbol{H}}} = \tilde{H}_x, \tilde{H}_y, \tilde{H}_z$ represent the electric and magnetic field vectors in three dimensions, respectively. The symbols μ, σ and ϵ denote the magnetic permeability, electrical conductivity, and dielectric permittivity, respectively. The variable ω stands for the angular frequency and i denotes the imaginary unit.

To emulate an infinite computational domain, we implement perfectly matched layer (PML) boundary conditions where we consider the general case of a complex stretched coordinate space [51]. In this scenario, the gradient operator ∇ takes the form

$$\nabla = \frac{\hat{x}}{p_x}\frac{\partial}{\partial x} + \frac{\hat{y}}{p_y}\frac{\partial}{\partial y} + \frac{\hat{z}}{p_z}\frac{\partial}{\partial z}, \tag{3}$$

where

$$p_k = \kappa_k + \frac{\sigma_k}{\alpha_k + i\omega\varepsilon_0}, \quad k = x, y, z \tag{4}$$

is a complex coordinate stretching variable that varies solely in the k-direction. Variable ε_0 represents the dielectric permittivity of free space. In the PML region, $\kappa_k > 1$ and $\sigma_k > 0$. In the non-PML region, $\kappa_k = 1$, $\sigma_k = 0$, and $\alpha_k = 0$, meaning that (3) reverts to the standard form of the gradient operator.

For 2D modelling, we assume no variation in the y-dimension. As a result, equations (1) and (2) can be split into two independent sets of equations referred to as the transverse-magnetic (TM) and transverse-electric (TE) modes. Here, our focus is on determining the explicit matrix-based expression of the TE-mode equations, which are given by

$$i\omega\mu\tilde{H}_x = -\frac{1}{p_z}\frac{\partial\tilde{E}_y}{\partial z}, \tag{5}$$

$$i\omega\mu\tilde{H}_z = \frac{1}{p_x}\frac{\partial\tilde{E}_y}{\partial x}, \tag{6}$$

$$\sigma\tilde{E}_y + i\omega\varepsilon\tilde{E}_y = \frac{1}{p_x}\frac{\partial\tilde{H}_z}{\partial x} - \frac{1}{p_z}\frac{\partial\tilde{H}_x}{\partial z}. \tag{7}$$

The derivation of the TM-mode expressions is similar. For EM wave propagation simulations over large model domains, the finite-difference time-domain (FDTD) method outperforms the finite-difference frequency-domain (FDFD) method in terms of computational memory efficiency. The reason for this is that the latter requires the construction of a Helmholtz matrix, the size of which increases quadratically with the size of model, whereas the former only demands allocating three snapshots in memory. We thus move from the frequency domain to the time domain by taking the inverse Fourier transform of (5) through (7). The convolutional perfectly matched layer (CPML) method is employed to prevent the virtual split of electric and magnetic field components [52]. Following the derivation in [14], the time-domain expression of TE-mode equations with CPML boundary conditions is

$$\mu\frac{\partial H_x}{\partial t} = -\frac{1}{\kappa_z}\frac{\partial E_y}{\partial z} - \zeta_z * \frac{\partial E_y}{\partial z}, \tag{8}$$

$$\mu\frac{\partial H_z}{\partial t} = \frac{1}{\kappa_x}\frac{\partial E_y}{\partial x} + \zeta_x * \frac{\partial E_y}{\partial x}, \tag{9}$$

$$\begin{aligned}\sigma E_y + \varepsilon\frac{\partial E_y}{\partial t} =& \frac{1}{\kappa_x}\frac{\partial H_z}{\partial x} - \frac{1}{\kappa_z}\frac{\partial H_x}{\partial z} \\ &+ \zeta_x * \frac{\partial H_z}{\partial x} - \zeta_z * \frac{\partial H_x}{\partial z},\end{aligned} \tag{10}$$

where

$$\zeta_k(t) = -\frac{\sigma_k}{\varepsilon_0\kappa_k^2}exp[-\frac{t}{\varepsilon_0}(\frac{\sigma_k}{\kappa_k} + \alpha_k)]u, \quad k = x, y, z. \tag{11}$$

where $u = 0$ in the non-PML region and $u = 1$ within the PML.

III. Born approximation operator in matrix form

LSRTM requires a linear relationship between defined image and data vectors. This linear operator can be derived using the Born approximation, which is also employed in gradient-based full waveform inversion (FWI) algorithms to establish a linear connection between model and data perturbations [54], assuming single-order scattering model. LSRTM utilizes the same operator for its forward problem, but it differentiates itself from FWI by substituting the data perturbation with

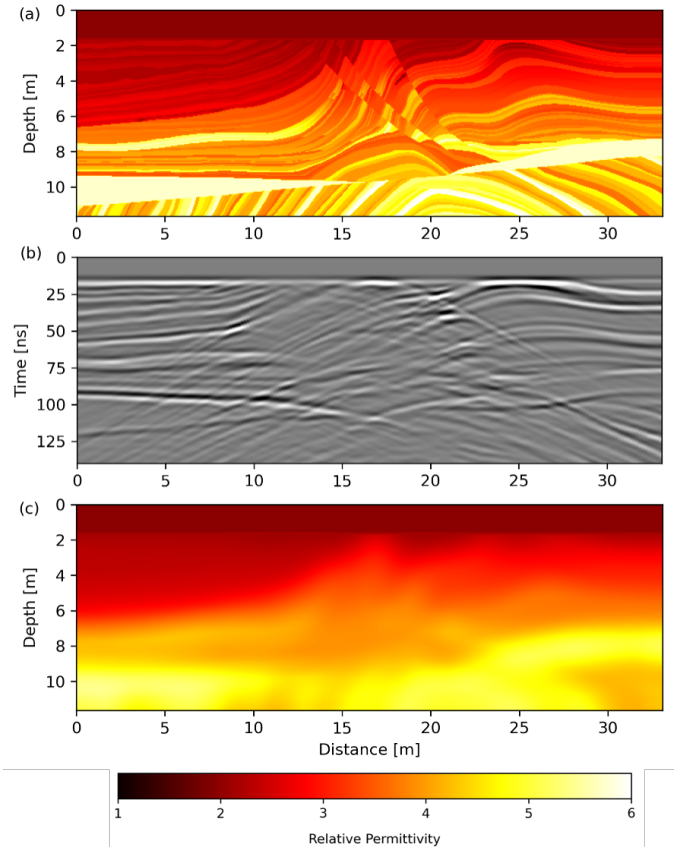

Fig. 1. Model representations and data: (a) Relative dielectric permittivity model adapted from the well-established Marmousi model used for benchmarking seismic migration algorithms [53], (b) corresponding zero-offset GPR reflection data, and (c) background permittivity model employed for migration.

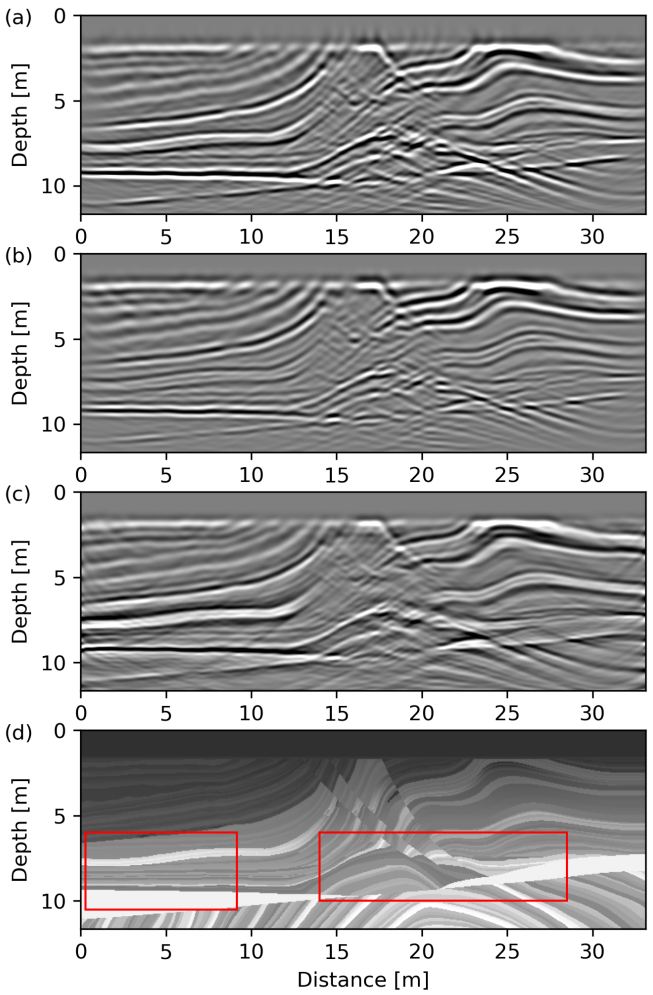

Fig. 2. Comparison of RTM results obtained using (a) post-stack migration, (b) pre-stack migration, and (c) LSRTM, along with (d) the true permittivity model from Fig. 1a for reference. The two regions delineated by red boxes in (d) are shown in detail in Figs. 3 and 4. The true permittivity model is displayed in grayscale for ease of comparison.

the primary reflections, which, in turn, results in an image of the reflections. The physical interpretation of the Born approximation operator is explored in greater detail in Section IV.

We focus our development here on the non-PML region of the simulation domain because the structural features within the PML region are not of interest and will not be subject to iterative updates in the LSRTM algorithm. In the non-PML region ($\zeta_k(t) = 0$ and $\kappa_k = 1$), we can simplify (8) through (10) as follows:

$$\mu\frac{\partial H_x}{\partial t} + \frac{\partial E_y}{\partial z} = s_1, \tag{12}$$

$$\mu\frac{\partial H_z}{\partial t} - \frac{\partial E_y}{\partial x} = s_2, \tag{13}$$

$$\sigma E_y + \varepsilon\frac{\partial E_y}{\partial t} - \frac{\partial H_z}{\partial x} + \frac{\partial H_x}{\partial z} = s_3, \tag{14}$$

where s_1, s_2, and s_3 denote source functions associated with the wavefields H_x, H_z, and E_y, respectively.

Let us express electrical properties μ, σ, and ε as the sum of a background large-scale model and local perturbation, $\mu = \mu_b + \Delta\mu$, $\sigma = \sigma_b + \Delta\sigma$, and $\varepsilon = \varepsilon_b + \Delta\varepsilon$, where b and Δ stand for the background model and perturbation respectively. Similarly, the corresponding electric and magnetic field components can be decomposed as $H_x = H_{xb} + \Delta H_x$, $H_z = H_{zb} + \Delta H_z$, and $E_y = E_{yb} + \Delta E_y$. Substituting these decompositions into (12) to (14) and assuming that multiplications between model perturbations and field perturbations (e.g., $\Delta\mu\frac{\partial\Delta H_x}{\partial t}$, $\Delta\sigma\Delta E_y$) are sufficiently small in the vicinity of the model perturbations such that they can be neglected, we arrive at two sets of equations. The first is

$$\mu_b\frac{\partial H_{xb}}{\partial t} + \frac{\partial E_{yb}}{\partial z} = s_1, \tag{15}$$

$$\mu_b\frac{\partial H_{zb}}{\partial t} - \frac{\partial E_{yb}}{\partial x} = s_2, \tag{16}$$

$$\sigma_b E_{yb} + \varepsilon_b\frac{\partial E_{yb}}{\partial t} - \frac{\partial H_{zb}}{\partial x} + \frac{\partial H_{xb}}{\partial z} = s_3, \tag{17}$$

where H_{xb}, H_{zb}, and E_{yb} represent source wavefields that are excited by the source functions s_1, s_2, and s_3. For the purpose of migration, as the background medium and source function are provided, these source wavefields can be calculated in advance and remain constant.

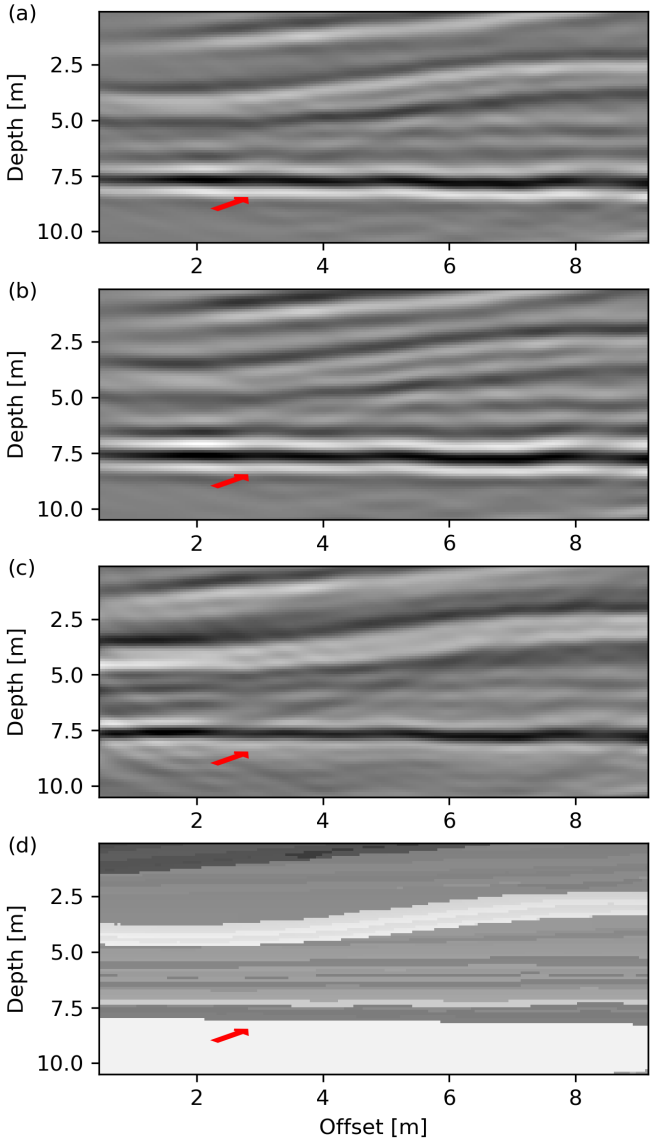

Fig. 3. Zoom of the left-hand box in Fig. 2 to better illustrate the structures resolved by (a) post-stack RTM, (b) pre-stack RTM, (c) LSRTM. The true permittivity modelis presented in grayscale in (d) for ease of comparison. The red arrows in each figure indicate the same subsurface location, highlighting the position of the side-lobes.

The second set of equations is

$$\mu_b \frac{\partial \Delta H_x}{\partial t} + \frac{\partial \Delta E_y}{\partial z} = r_1, \tag{18}$$

$$\mu_b \frac{\partial \Delta H_z}{\partial t} - \frac{\partial \Delta E_y}{\partial x} = r_2, \tag{19}$$

$$\sigma_b \Delta E_y + \varepsilon_b \frac{\partial \Delta E_y}{\partial t} - \frac{\partial \Delta H_z}{\partial x} + \frac{\partial \Delta H_x}{\partial z} = r_3, \tag{20}$$

where

$$r_1 = -\Delta\mu \frac{\partial H_{xb}}{\partial t}, \tag{21}$$

$$r_2 = -\Delta\mu \frac{\partial H_{zb}}{\partial t}, \tag{22}$$

$$r_3 = -\Delta\sigma E_{yb} - \Delta\varepsilon \frac{\partial E_{yb}}{\partial t}. \tag{23}$$

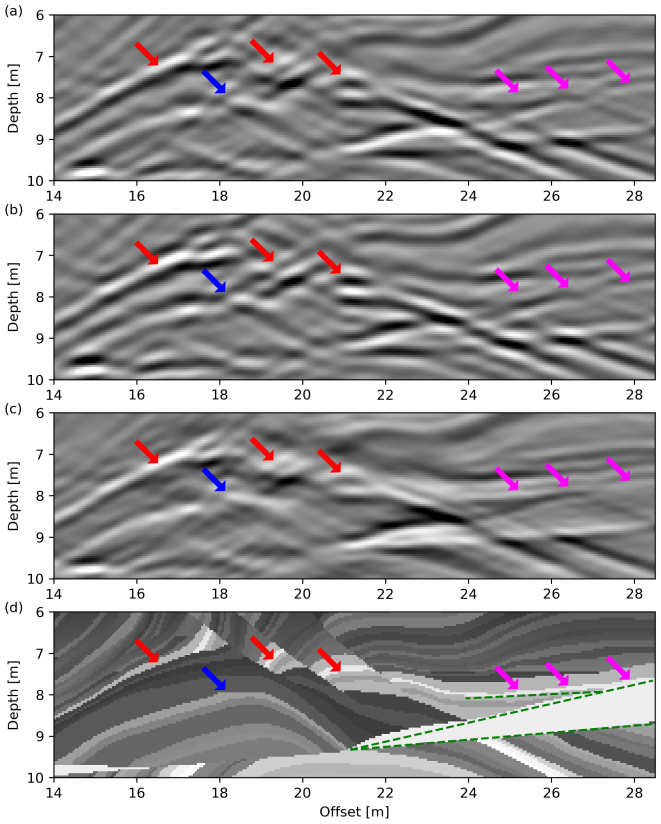

Fig. 4. Zoom of the right-hand box in Fig. 2 to better illustrate the structures resolved by (a) post-stack RTM, (b) pre-stack RTM, (c) LSRTM. The true permittivity modelis presented in grayscale in (d) for ease of comparison. The green dashed lines are not drawn on the migrated images because they hinder the interpretation.

Equations (21) through (23) can be interpreted as the subsurface responses (r_1, r_2, and r_3) to the modified source wavefields ($\frac{\partial H_{xb}}{\partial t}$, $\frac{\partial H_{zb}}{\partial t}$, E_{yb}, and $\frac{\partial E_{yb}}{\partial t}$) at every model perturbation location ($\Delta\mu$, $\Delta\sigma$, and $\Delta\varepsilon$) and across all time samples.

To represent (18) to (23) in matrix form, we discretize the model domain into nm elements and the time axis into nt temporal samples. Then, each wavefield has $nm \times nt$ elements. Equations (21) through (23) can then be expressed as

$$\mathbf{r} = \mathbf{K}\mathbf{S}\Delta\mathbf{m}, \tag{24}$$

where $\Delta\mathbf{m} = [\Delta\mu, \Delta\sigma, \Delta\varepsilon]$ is a column vector of size $3nm \times 1$ and $\mathbf{S}$ is a spray operator in time [55] that duplicates the model for each time sample and whose size is $3nm \cdot nt \times 3nm$. Matrix $\mathbf{K}$ has the form

$$\mathbf{K} = \begin{bmatrix} -\frac{\partial H_{xb}}{\partial t} & \mathbf{0} & \mathbf{0} \\ -\frac{\partial H_{zb}}{\partial t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -E_{yb} & -\frac{\partial E_{yb}}{\partial t} \end{bmatrix}, \tag{25}$$

where the modified source wavefields, $\frac{\partial H_{xb}}{\partial t}$, $\frac{\partial H_{zb}}{\partial t}$, E_{yb}, and $\frac{\partial E_{yb}}{\partial t}$ are arranged into diagonal sub-matrices having a size of $nm \cdot nt \times nm \cdot nt$. As mentioned above, these wavefields can be computed beforehand based on the provided background medium and source function and thus are assumed to be known. Hence, $\mathbf{K}$ is constant in LSRTM.

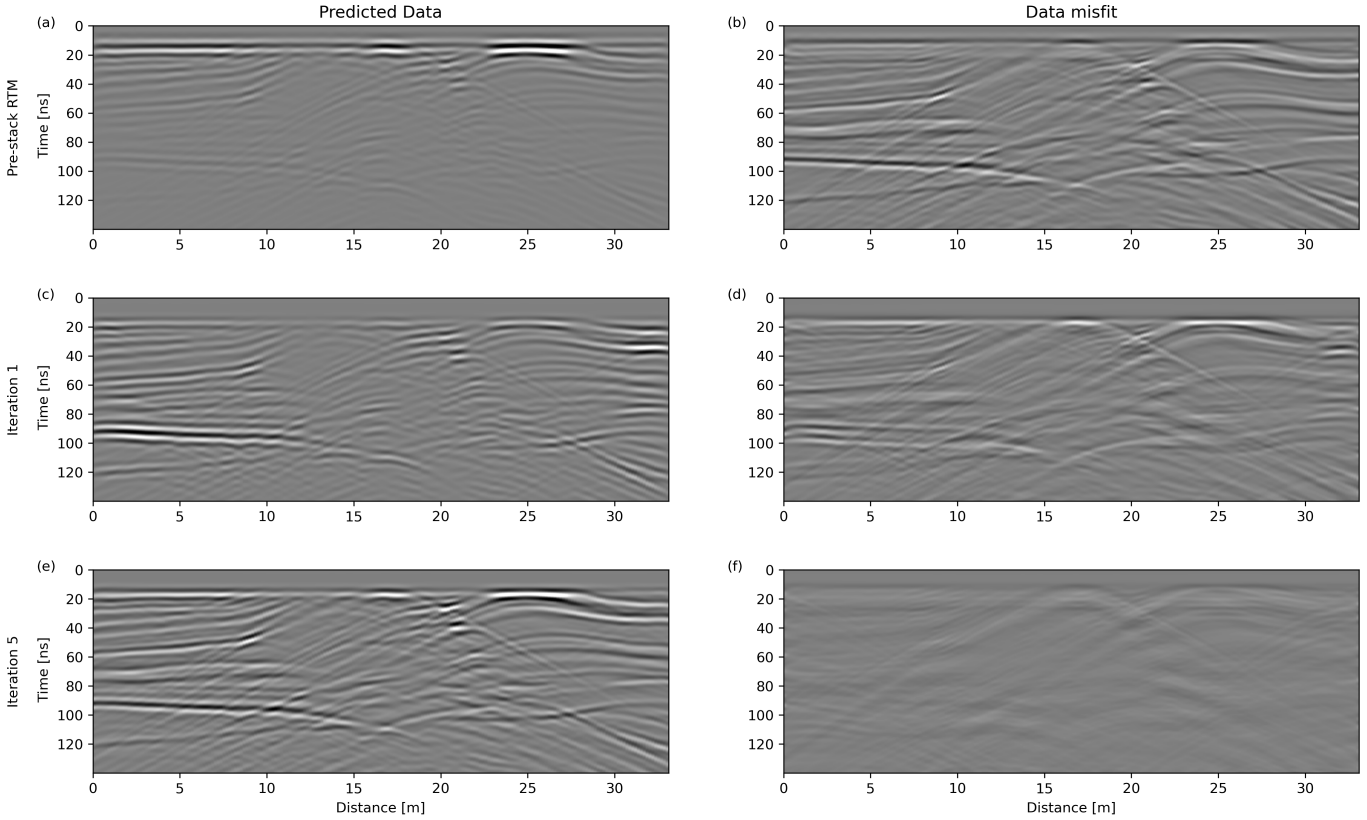

Fig. 5. GPR data predicted by the LSRTM forward operator using (a) the image obtained from pre-stack RTM, (c) the image obtained after 1 iteration of LSRTM, and (e) the image obtained after 5 iterations of LSRTM. The corresponding differences with the observed data are shown in (b), (d) and (f), respectively.

The form of (18) through (20) is similar to that of (15) through (17). The primary distinction is the source function term. Equations (18) through (20) suggest that the perturbed wavefields ΔH_x, ΔH_z and ΔE_y are excited by the subsurface responses $\mathbf{r} = [r_1, r_2, r_3]$. These equations can also be expressed in matrix form as

$$\Delta \mathbf{w} = \mathbf{G}\mathbf{r}, \tag{26}$$

where $\mathbf{G}$ represents the FDTD simulation of TE-mode equations and is of size $3nm \cdot nt \times 3nm \cdot nt$. The matrix representation of $\mathbf{G}$ is non-trivial and will be discussed in detail in section V. This operator takes the subsurface response $\mathbf{r}$ as its source function to generate the perturbed wavefield vector $\Delta \mathbf{w} = [\Delta \mathbf{H}_x, \Delta \mathbf{H}_z, \Delta \mathbf{E}_y]$.

Lastly, we introduce a sampling operator $\mathbf{R}$ to extract the wavefield at receiver antenna locations along the surface for all time samples. Then, we have

$$\Delta \mathbf{d} = \mathbf{R}\Delta \mathbf{w}. \tag{27}$$

The matrix dimension of $\mathbf{R}$ is $nd \times 3 \cdot nm$, where nd denotes the total number of data samples for all three field components ($\Delta \mathbf{H}_x$, $\Delta \mathbf{H}_z$, $\Delta \mathbf{E}_y$). The resulting vector $\Delta \mathbf{d}$ contains the data perturbation in the form of a column vector. Note that, in practice, most GPR receiver antennas only measure a single component of the electrical field, such as $\mathbf{E}_y$. In this case, the sampling operator should only extract information on that particular component, which is equivalent to removing the rows related to components H_x and H_z in matrix $\mathbf{R}$.

Combining (24), (26) and (27), a linear relationship between model perturbations and data perturbations can be established as follows:

$$\Delta \mathbf{d} = \mathbf{RGKS}\Delta \mathbf{m} \tag{28}$$

$$= \mathbf{L}\Delta \mathbf{m}, \tag{29}$$

where $\mathbf{L}$ denotes the Born approximation operator, which is also known as the *demigration* operator [56]. Note that the matrices in (28) are used solely for the mathematical derivation. In our computer code, they are not formed explicitly, but rather designed as computational operators for the purpose of memory efficiency.

IV. Least-squares reverse time migration

Understanding the physical meaning of $\Delta \mathbf{m}$ and $\Delta \mathbf{d}$ is essential to the problem of LSRTM. The linear relationship $\Delta \mathbf{d} = \mathbf{L}\Delta \mathbf{m}$ is also a concept that can be found in the field of FWI [54] when is solved by gradient-based optimization algorithms, and it is important to clarify the differences in interpretation between these two applications. In FWI, $\Delta \mathbf{d}$ represents the misfit between the predicted GPR data corresponding to the current electrical property model and the actual recorded data. Hence, $\Delta \mathbf{m}$ in this scenario reflects the deviation between the current model and the sought electrical

property model that can optimally predict all waveforms. In LSRTM, on the other hand, $\Delta\mathbf{d}$ represents primary reflections. The model $\Delta\mathbf{m}$ in this case represents the corresponding reflectors.

To gain further clarity, let us consider that the source wavefields $\mathbf{K}$ in LSRTM can be regarded as incident waves descending into the subsurface. These wavefields are related to the background model and we can assume that the smoothness of this model mainly supports downward propagation with minimal backscattering. Meanwhile, we assume that the reflections are mainly created by the perturbed model $\Delta\mathbf{m}$. Hence, vector $\mathbf{r}$ in (24) can be conceptualized as scattering strength, and the combined application of (26) and (27) simulates the journey of the scattered energy to the receiver antennas at the surface. Thus, the operator $\mathbf{L}$ effectively approximates the physical process of primary reflection in a linear manner. When $\Delta\mathbf{d}$ is regarded as the recorded reflections in the LSRTM problem, $\Delta\mathbf{m}$ should therefore represent a model of the reflectors.

Based on the above, we hereby replace $\Delta\mathbf{d}$ by the observed GPR reflection data, $\mathbf{d}^{obs}$, and we denote the corresponding migrated reflection image as $\mathbf{m}$. Depending on the quality of the observed data $\mathbf{d}^{obs}$, pre-processing may be necessary. This may involve steps such as data denoising and amplitude gain adjustment. As LSRTM necessitates that $\mathbf{d}^{obs}$ consists of primary reflection data, removing multiples may also be required. In our experience, however, the latter are rare in GPR surveys and only occur with significant amplitudes in rather pathological settings.

The goal of LSRTM is to find the image that best reproduces the data. Mathematically, this is done by minimizing the following cost function with regard to $\mathbf{m}$:

$$J = ||\mathbf{L}\mathbf{m} - \mathbf{d}^{obs}||_2^2 + \mu^2||\mathbf{W}_m\mathbf{m}||_2^2, \tag{30}$$

where the weighting operator $\mathbf{W}_m$ is designed to be independent of $\mathbf{m}$ to maintain linearity in the cost function. In its simplest form, $\mathbf{W}_m$ can be an identity matrix $\mathbf{I}$. Alternatively, $\mathbf{W}_m$ may serve as a filter to eliminate artifacts in the image. For this study, we adopt the approach from [47] to design $\mathbf{W}_m$ for normalizing the spatial sensitivity in the image. To this end, it is structured as a diagonal matrix, with its diagonal elements given by

$$diag(\mathbf{W}_m) = \sum_t (\mathbf{G}\mathbf{s})^2. \tag{31}$$

This ensures that the optimized image has more uniform amplitudes compared to one produced with $\mathbf{W}_m = \mathbf{I}$. Furthermore, as we will demonstrate in Section VIII-A, this method of sensitivity normalization through $\mathbf{W}_m$ can also help to accelerate the convergence rate.

Parameter μ^2 in (30) controls the balance between data misfit and model regularization in the LSRTM procedure. In principle, multiple values of μ^2 should be tested to find the optimum trade-off between these two objectives. However, this approach is computationally too expensive for optimization problems such as LSRTM. A practical solution is to adopt an iterative method to optimize $\mathbf{m}$ and to use the number of iterations to implicitly control the damping parameter [57]. To this end, we transform the regularized least-squares inverse problem (Eq. (30)) into the damped least-squares problem

$$J = \frac{1}{2}||\mathbf{L}\mathbf{W}_m^{-1}\mathbf{u} - \mathbf{d}^{obs}||_2^2 + \frac{1}{2}\mu^2||\mathbf{u}||_2^2, \tag{32}$$

where

$$\mathbf{m} = \mathbf{W}_m^{-1}\mathbf{u} \tag{33}$$

and $\mathbf{u}$ serves as an auxiliary variable to be optimized. In this form, employing a larger number of iterations results in a migrated image with a lower data misfit but with less regularization. Conversely, using fewer iterations can help to avoid the problem of over-fitting. Note that (33) is equivalent to the use of the illumination-compensated imaging condition [58].

The LSRTM problem can be efficiently solved using the conjugate gradients algorithm [59], [60], which requires the forward operator $\mathbf{L}$ and its adjoint. Obtaining the adjoint operator is a critical topic for all gradient-based optimization methods. Given that $\mathbf{L}$ in (29) is represented by a real-valued matrix, the adjoint operator is simply equal to the transpose

$$\mathbf{L}^T = \mathbf{S}^T\mathbf{K}^T\mathbf{G}^T\mathbf{R}^T, \tag{34}$$

where $\mathbf{R}^T$ can be interpreted as an operator that injects data into an initial wavefield containing zeros at each receiver antenna location and at all times. If, as is normally the case, only component $\mathbf{E}_y$ is recorded as data, then $\mathbf{R}^T$ should inject zeros into $\mathbf{H}_x$ and $\mathbf{H}_z$. Matrix $\mathbf{G}^T$ is called the adjoint simulation operator, and is described in detail in Section VI. Matrix $\mathbf{K}^T$ can be easily obtained as its non-zero sub-matrices are diagonal. Matrix $\mathbf{S}^T$ is an integration operator over time.

V. Discrete Simulation Operator: Matrix Expression of the Finite-Difference Time-Domain Method

In this section, we formulate the simulation operator for Maxwell's TE-mode equations using matrices. To this end, the FDTD method is utilized to solve for the electric and magnetic field components in (8) through (10).

First, we discretize the subsurface distribution of the electrical properties using a staggered grid [61]. A higher-order finite-difference [62] is adopted to approximate the spatial derivatives of the field components, whereas the temporal derivatives are approximated by a second-order-accurate stencil. The convolution terms are computed using a recursive scheme [63]. The evolution of both fields in time is accomplished through a leap-frog approach.

The expression for the finite-difference update of the aforementioned equations is

$$\begin{aligned} H_x|_{ix,iz+0.5}^{it+0.5} &= H_x|_{ix,iz+0.5}^{it-0.5} \\ &- D_c|_{ix,iz+0.5}\Psi_{Hxz}|_{ix,iz+0.5}^{it} \\ &- D_{bz}|_{ix,iz+0.5}\sum_1^N c_n(E_y|_{ix,iz+n}^{it} - E_y|_{ix,iz-n+1}^{it}), \end{aligned} \tag{35}$$

$$\begin{aligned} H_z|_{ix+0.5,iz}^{it+0.5} &= H_z|_{ix+0.5,iz}^{it-0.5} \\ &- D_c|_{ix+0.5,iz}\Psi_{Hzx}|_{ix+0.5,iz}^{it} \\ &- D_{bx}|_{ix+0.5,iz}\sum_1^N c_n(E_y|_{ix+n,iz}^{it} - E_y|_{ix-n+1,iz}^{it}), \end{aligned} \tag{36}$$

$$\begin{aligned} E_y|_{ix,iz}^{it+1} &= C_a|_{ix,iz}E_y|_{ix,iz}^{it} \\ &- C_c|_{ix,iz}(\Psi_{Eyx}|_{ix,iz}^{it+0.5} - \Psi_{Eyz}|_{ix,iz}^{it+0.5}) \\ &+ C_{bx}|_{ix,iz}\sum_1^N c_n(H_z|_{ix+n-0.5,iz}^{it+0.5} - H_z|_{ix-n+0.5,iz}^{it+0.5}) \\ &- C_{bz}|_{ix,iz}\sum_1^N c_n(H_x|_{ix,iz+n-0.5}^{it+0.5} - H_x|_{ix,iz-n+0.5}^{it+0.5}), \end{aligned} \tag{37}$$

where c_n denotes the n^{th} order finite-difference coefficients in space, and it, ix and iz denote temporal, x-axis and z-axis indices, respectively. The coefficients C_a, C_{bk}, C_c, D_{bk} and D_c are specified as

$$C_a = (1 - \frac{\sigma\Delta t}{2\varepsilon})(1 + \frac{\sigma\Delta t}{2\varepsilon})^{-1}, \tag{38}$$

$$C_{bk} = \frac{\Delta t}{\varepsilon}(1 + \frac{\sigma\Delta t}{2\varepsilon})^{-1}(\kappa_k\Delta k)^{-1}, \tag{39}$$

$$C_c = \frac{\Delta t}{\varepsilon}(1 + \frac{\sigma\Delta t}{2\varepsilon})^{-1}, \tag{40}$$

$$D_{bk} = \frac{\Delta t}{\mu}(\kappa_k\Delta k)^{-1}, \tag{41}$$

$$D_c = \frac{\Delta t}{\mu}, \tag{42}$$

where $k = x, z$. The PML memory variables $\Psi_{(.)}$ are given by

$$\begin{aligned} \Psi_{Hxz}|_{ix,iz+0.5}^{it} &= B_z|_{ix,iz+0.5}\Psi_{Hxz}|_{ix,iz+0.5}^{it-1} \\ &+ A_z|_{ix,iz+0.5}\sum_1^N c_n(E_y|_{ix,iz+n}^{it} - E_y|_{ix,iz-n+1}^{it}), \end{aligned} \tag{43}$$

$$\begin{aligned} \Psi_{Hzx}|_{ix+0.5,iz}^{it} &= B_x|_{ix+0.5,iz}\Psi_{Hzx}|_{ix+0.5,iz}^{it-1} \\ &+ A_x|_{ix+0.5,iz}\sum_1^N c_n(E_y|_{ix+n,iz}^{it} - E_y|_{ix-n+1,iz}^{it}), \end{aligned} \tag{44}$$

$$\begin{aligned} \Psi_{Eyx}|_{ix,iz}^{it+0.5} &= B_x|_{ix,iz}\Psi_{Eyx}|_{ix,iz}^{it-0.5} \\ &+ A_x|_{ix,iz}\sum_1^N c_n(H_z|_{ix+n-0.5,iz}^{it+0.5} - H_z|_{ix-n+0.5,iz}^{it+0.5}), \end{aligned} \tag{45}$$

$$\begin{aligned} \Psi_{Eyz}|_{ix,iz}^{it+0.5} &= B_z|_{ix,iz}\Psi_{Eyz}|_{ix,iz}^{it-0.5} \\ &+ A_z|_{ix,iz}\sum_1^N c_n(H_x|_{ix,iz+n-0.5}^{it+0.5} - H_x|_{ix,iz-n+0.5}^{it+0.5}), \end{aligned} \tag{46}$$

where

$$A_k = \frac{\sigma_k}{\sigma_k\kappa_k + \alpha_k\kappa_k^2}(B_k - 1)\Delta k^{-1}, \tag{47}$$

$$B_k = exp[-\frac{\Delta t}{\varepsilon_0}(\frac{\sigma_k}{\kappa_k} + \alpha_k)]. \tag{48}$$

Note that, in the above equations, the multiplications between the coefficients A, B, C and D and the field components are performed element-wise. If the field components are reshaped into column vectors, these multiplications can be expressed as a diagonal matrix multiplying a vector. This allows the finite-differences to be expressed in the form of matrix-vector multiplications. Consequently, the updates of the electric and magnetic fields in (35) to (37) can be fully represented in matrix form as

$$\mathbf{H}_x^{n+1} = \mathbf{H}_x^n - \mathbf{D}_{bz}\mathbf{F}_{zf}\mathbf{E}_y^n - \mathbf{D}_c\mathbf{\Psi}_{Hxz}^n \tag{49}$$

$$\mathbf{H}_z^{n+1} = \mathbf{H}_z^n + \mathbf{D}_{bx}\mathbf{F}_{xf}\mathbf{E}_y^n + \mathbf{D}_c\mathbf{\Psi}_{Hzx}^n \tag{50}$$

$$\begin{aligned} \mathbf{E}_y^{n+1} =& \mathbf{C}_a\mathbf{E}_y^n + \mathbf{C}_{bx}\mathbf{F}_{xb}\mathbf{H}_z - \mathbf{C}_{bz}\mathbf{F}_{zb}\mathbf{H}_x \\ &+ \mathbf{C}_c[\mathbf{\Psi}_{Eyx}^n - \mathbf{\Psi}_{Eyz}^n], \end{aligned} \tag{51}$$

where $\mathbf{E}_k$ and $\mathbf{H}_k$ are column vectors, and $\mathbf{D}_{(.)}$ and $\mathbf{C}_{(.)}$ are diagonal matrices. Likewise, the updates of $\mathbf{\Psi}_{(.)}$ can also be expressed as

$$\mathbf{\Psi}_{Hxz}^n = \mathbf{B}_z\mathbf{\Psi}_{Hxz}^{n-1} + \mathbf{A}_z\mathbf{F}_{zf}\mathbf{E}_y^n, \tag{52}$$

$$\mathbf{\Psi}_{Hzx}^n = \mathbf{B}_x\mathbf{\Psi}_{Hzx}^{n-1} + \mathbf{A}_x\mathbf{F}_{xf}\mathbf{E}_y^n, \tag{53}$$

$$\mathbf{\Psi}_{Eyx}^n = \mathbf{B}_x\mathbf{\Psi}_{Eyx}^n + \mathbf{A}_x\mathbf{F}_{xb}\mathbf{H}_z, \tag{54}$$

$$\mathbf{\Psi}_{Eyz}^n = \mathbf{B}_z\mathbf{\Psi}_{Eyz}^n + \mathbf{A}_z\mathbf{F}_{zb}\mathbf{H}_x. \tag{55}$$

Matrices $\mathbf{F}_{zf}$ and $\mathbf{F}_{zb}$ are finite-difference operators along the z-direction with front-end and back-end zero-padding, respectively. Given that the field components are situated on a staggered grid, the zero-padding importantly serves to adjust the center of the finite-difference stencil in space. Similarly, $\mathbf{F}_{xf}$ and $\mathbf{F}_{xb}$ are finite-difference operators along the x-direction, also with front-end and back-end zero-padding.

Note that n in (49) through (55) does not denote the n^{th} time sample. Instead, it indicates the n^{th} update of the field. These equations are not solved simultaneously because $\mathbf{E}_{(k)}$, $\mathbf{H}_{(k)}$, and $\mathbf{\Psi}_{(k)}$ are distributed along a staggered time axis. For instance, if the simulation starts with $E|^{it=0}ix, iz$ as the known variable, the finite-difference update at the n^{th} iteration can be summarized by

1) Update $\mathbf{\Psi}_{Hxz}$ and $\mathbf{\Psi}_{Hzx}$ by solving (52) and (53),
2) Update $\mathbf{H}_x$ and $\mathbf{H}_z$ by solving (49) and (50),
3) Update $\mathbf{\Psi}_{Eyz}$ and $\mathbf{\Psi}_{Eyx}$ by solving (55) and (54),
4) Update $\mathbf{E}_y$ by solving (51).

Alternatively, we can express a single-step finite-difference update from iteration n to $n+1$ in matrix form as

$$\mathbf{w}^{n+1} = \mathbf{T}\mathbf{w}^n + \mathbf{s}^{n+1}, \tag{56}$$

where $\mathbf{w}^{(.)} = [\mathbf{H}_x, \mathbf{H}_z, \mathbf{E}_y, \mathbf{\Psi}_{Hxz}, \mathbf{\Psi}_{Hzx}, \mathbf{\Psi}_{Eyx}, \mathbf{\Psi}_{Eyz}]$ is a column vector that contains the wavefield components, and $\mathbf{s}$ is a zero vector having the same dimensions as $\mathbf{w}$. The source function $r_{(.)}$ is assigned at each source location for all time samples. Matrix $\mathbf{T} = \mathbf{T}_4\mathbf{T}_3\mathbf{T}_2\mathbf{T}_1$ is the finite-difference time-stepping operator whose expression is provided in Appendix A.

If the initial condition of the wavefield is given as $\mathbf{w}^0 = \mathbf{s}^0 = \mathbf{0}$, then for $n = 1, 2, 3$, equation (56) becomes

$$\mathbf{w}^1 = \mathbf{T}\mathbf{w}^0 + \mathbf{s}^1 = \mathbf{s}^1 \tag{57}$$

$$\mathbf{w}^2 = \mathbf{T}\mathbf{w}^1 + \mathbf{s}^2 \tag{58}$$

$$\mathbf{w}^3 = \mathbf{T}\mathbf{w}^2 + \mathbf{s}^3. \tag{59}$$

This is equivalent to writing

$$\mathbf{w}^1 = \mathbf{s}^1, \tag{60}$$

$$\mathbf{w}^2 = \mathbf{T}\mathbf{s}^1 + \mathbf{s}^2, \tag{61}$$

$$\mathbf{w}^3 = \mathbf{T}^2\mathbf{s}^1 + \mathbf{T}\mathbf{s}^2 + \mathbf{s}^3. \tag{62}$$

Equations (60) to (62) can be written in matrix form as

$$\begin{bmatrix} \mathbf{w}^1 \\ \mathbf{w}^2 \\ \mathbf{w}^3 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{T} & \mathbf{I} & \mathbf{0} \\ \mathbf{T}^2 & \mathbf{T} & \mathbf{I} \end{bmatrix}}_{\mathbf{G}} \begin{bmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \mathbf{s}^3 \end{bmatrix}. \tag{63}$$

In this case, for these three updates only, equation (63) requires four matrix-vector multiplications with $\mathbf{T}$. In the general case, computing n wavefield updates requires $\frac{n(n-1)}{2}$ multiplications with $\mathbf{T}$.

Alternatively, equation (63) can be rewritten as

$$\begin{bmatrix} \mathbf{w}^1 \\ \mathbf{w}^2 \\ \mathbf{w}^3 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{T} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}}_{\mathbf{G}} \begin{bmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \mathbf{s}^3 \end{bmatrix}. \tag{64}$$

In this form, computing three wavefield updates requires only two multiplications with $\mathbf{T}$, and more generally, computing n updates requires merely $n-1$ multiplications. Equation (64) is identical to the expression of acoustic FDTD simulation [50]. Intrinsically, it is the general expression of the temporal second-order finite-difference method in matrix form.

VI. Adjoint Simulation Operator: Exact Adjoint versus Pseudo Adjoint

Identifying the Born approximation in matrix form followed by taking the matrix transpose (Eq. (34)) is known as the *discrete adjoint-state* method [64]. It leads to the formulation of the adjoint *simulation* operator as the matrix transpose of $\mathbf{G}$ in (64), which is given by

$$\mathbf{G}^T = \begin{bmatrix} \mathbf{I} & \mathbf{T}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{T}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{65}$$

Note, however, that the adjoint operator can also be obtained using the so-called *discretized adjoint-state* method (e.g., [35]), which employs Lagrange multipliers to derive an expression for the gradient in the continuous domain, followed by numerical discretization. Although both the discrete and discretized adjoint-state methods represent the same gradient expression, they carry out the adjoint simulation differently. Specifically, the discretized adjoint-state method uses time-reversed simulation. In matrix form, this is expressed as

$$\mathbf{G}^T_{pseudo} = \begin{bmatrix} \mathbf{I} & \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{T} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{66}$$

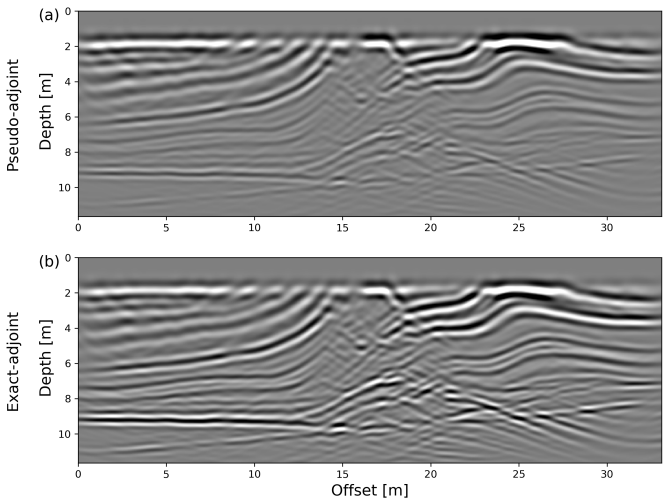

Fig. 6. Comparison of pre-stack RTM results obtained using (a) the pseudo adjoint simulation operator and (b) the exact adjoint simulation operator.

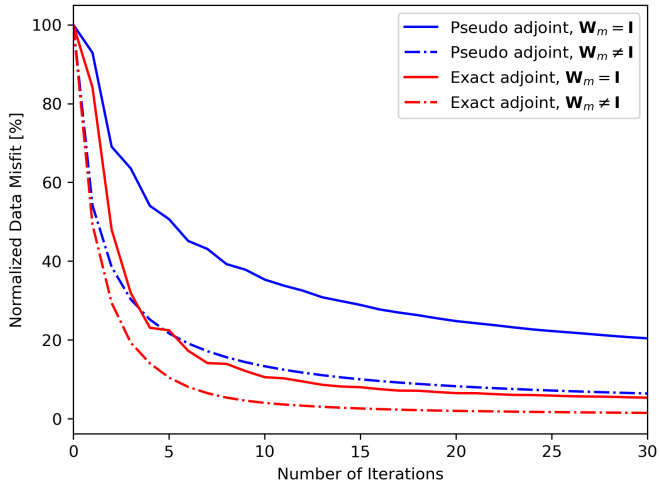

Fig. 7. Data misfit as a function of the number of LSTRM iterations, comparing use of the pseudo versus exact adjoint simulation operators as well as setting weighting matrix $\mathbf{W}_m = \mathbf{I}$ versus according to [65].

Compared with the discrete adjoint simulation operator $\mathbf{G}^T$ in (65), the discretized adjoint simulation operator $\mathbf{G}^T_{pseudo}$ does not take the transpose of $\mathbf{T}$. In Appendix A, we demonstrate that $\mathbf{T}$ is not a symmetric matrix. Hence, the time-reversed simulation $\mathbf{G}^T_{pseudo}$ is not the exact adjoint of the time-forward simulation $\mathbf{G}$. Use of the inexact adjoint can cause the conjugate gradient algorithm to converge suboptimally or even prevent it from converging. In view of this, RTM and LSRTM results obtained by both the exact and inexact adjoint operators are comprehensively compared in Section VIII-A.

VII. Numerical Examples

We use an adaptation of the extensively investigated Marmousi model [53], developed to benchmark seismic migration algorithms, to assess the effectiveness of our LSRTM approach. With its complex structure, this model is ideal for evaluating the improvement in resolution provided by our

method. In our study, we adapted it to a relative dielectric permittivity distribution by linearly scaling the values in the original model to a range between 1 and 6 (Fig. 1a). For our test, we assume that the electrical conductivity and relative magnetic permeability are constant throughout the model domain and equal to 0.01 S/m and 1, respectively.

The electrical properties were discretized on a 234×663 grid with a spatial sampling interval of 0.05 m. To generate synthetic zero-offset GPR data, we employed the FDTD method considering 200 measurements evenly spaced by 0.165 m along the surface. The spatial derivatives were approximated by an eighth-order operator, whereas the temporal derivatives were approximated by a second-order operator. A time discretization of 0.04 ns was used. For each transmitter location, 3500 time steps were evaluated, resulting in a total recording time of 140 ns per trace. Prior to migration, we applied a top-mute filter to remove the direct arrival. The resulting synthetic GPR reflection data are displayed in Fig. 1b.

An inadequate estimation of the velocity model can compromise the quality of the migrated image. Indeed, migration assumes that the background or low-wavenumber velocity model is reasonably well-known. For our synthetic test, the dielectric permittivity model assumed for migration is depicted in Fig. 1c. The model was obtained by twice convolving the original permittivity distribution with a local averaging filter of size 30×30 grid cells, in order to create a challenging test case for LSRTM. After the smoothing process, we see that detailed structural information in the original permittivity model (Fig. 1a) completely removed. In practice, the large-scale velocity structure of the probed subsurface region can be constrained through diffraction focusing analysis (e.g., [66]). Alternatively, one can acquire a small number of common mid-point (CMP) gathers along the survey line and interpolate the inferred 1D velocity-depth profiles (e.g., [67]–[69]).

Typically, GPR imaging aims to resolve the interfaces of the subsurface velocity model, which is primarily controlled by the dielectric permittivity. Therefore, our LSRTM focuses exclusively on inverting for the reflectivity image of the dielectric permittivity. Although the electrical conductivity can have a strong influence on GPR data, it largely impacts attenuation rather than travel time. Hence, the effects of conductivity on generating reflections are much less important than the dielectric permittivity. Theoretically, the magnetic permeability has the potential to influence GPR reflections. However, typically this parameter is equal to its free-space value in most non-magnetic materials considered in GPR.

Figures 2a, 2b, and 2c show the migrated images obtained using post-stack RTM based on the point-diffraction model [17], pre-stack RTM based on the exact adjoint operator in Eq. (34), and LSRTM based on (30) and the exact adjoint (34), respectively. As mentioned earlier, incorporating operator $\mathbf{W}_m$ in (30) can boost the amplitudes in areas of low sensitivity. To assure a fair comparison across the different RTM algorithms, the inverse operator $\mathbf{W}_m^{-1}$ was applied to both the post-stack and pre-stack RTM images using(33), where $\mathbf{u}$ is replaced by the image and thus $\mathbf{m}$ is the amplitude enhanced image. When presenting the migrated images, it is important to note that we have applied amplitude normalization, limiting the range between -1 and 1. To guarantee a fair comparison, all migrated images shown in the same figure adhere to the same colour scale.

We see in Fig. 2 that the major subsurface structures are effectively recovered by both the post-stack and pre-stack RTM methods. However, LSRTM provides superior model resolution. Reflecting interfaces appear to be crisper and better defined, fine-scale features are accentuated, and the illumination of deep structures is noticeably improved. For a more detailed comparison of image resolution, two particular regions have been selected, as outlined by the red boxes in Fig. 2d.

Figure 3 depicts the structures outlined by the left-hand box in Fig. 2d, which is primarily comprised of gently dipping layers. The key difference between the migration results lies in the size of the wavelet side lobes. As indicated by the red arrows, side-lobe suppression in the LSRTM image is significantly more effective compared to the post-stack and pre-stack RTM images. This comprehensive suppression of side lobes leads to a higher-resolution image. Although the enhancement in resolution may not be decisive in this structurally simple region, it is highly advantageous for complex areas like the region outlined by the right-hand box in Fig. 2d, which is presented in Fig. 4. Indeed, in this figure, the three high permittivity anomalies indicated by the red arrows are better focused by LSRTM. The ultra-thin layer, denoted by the magenta arrow, remains undetected in the images produced by post-stack and pre-stack RTM, but it is clearly discernible in the LSRTM image. The wedge-shaped feature, delineated by the green dashed lines, displays improved resolution along its interfaces. Although identifying the layers within the anticline, denoted by the blue arrow, proves challenging for all three imaging methods, LSRTM provides some noticeable improvements compared to the other approaches.

The considerable improvements observed with LSRTM over post-stack and pre-stack RTM can be predominantly attributed to the data-fitting term in the cost function (30). Indeed, for the presented example, pre-stack RTM exhibits a 29.7% data misfit, whereas LSRTM has a misfit of only 1.2% after ten iterations of the conjugate gradients method. Figure 5 shows the predicted data obtained by applying the LSRTM forward operator to the migrated image from pre-stack RTM, and to the images from 1 and 5 iterations of LSRTM, along with the corresponding data misfit. In Figs. 5a and 5b, we see that the data predicted by pre-stack RTM are overly smooth, and that significant high-frequency information remains in the data misfit. Additionally, we see that pre-stack RTM mainly focuses the data at early times. This is because the Hessian matrix, $\mathbf{L}^T\mathbf{L}$, has much greater sensitivity to the shallow and large-scale subsurface structures [70], [71]. Figures 5c and 5d, on the other hand, show that, even after a single iteration, the LSRTM image can adequately predict the later time data. After five iterations, the LSRTM image faithfully predicts the data throughout early and late times (Figs. 5e and 5f), thereby yielding a higher-resolution image and better illumination enhancement of deep structures.

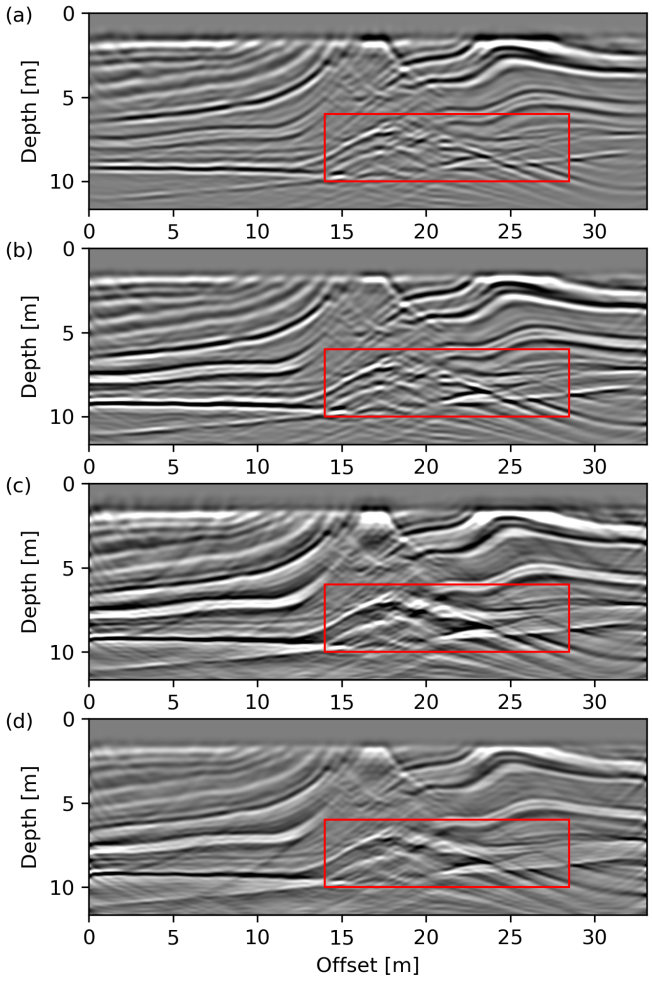

Fig. 8. Comparison of LSRTM results: (a) LSRTM with pseudo adjoint, (b) LSRTM with pseudo adjoint and regularization, (c) LSRTM with exact adjoint, and (d) LSRTM with exact adjoint and regularization.

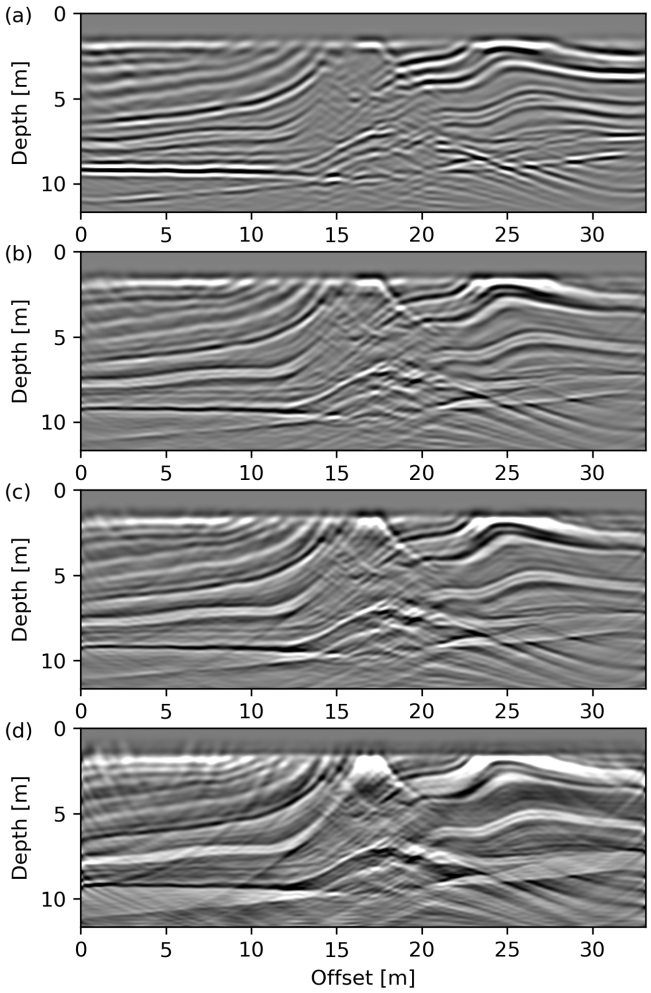

Fig. 9. Comparison of LSRTM results with the exact adjoint operator and regularization after (a) iteration 1, (b) iteration 5, (c) iteration 10, and (d) iteration 30.

VIII. Discussion

A. Pseudo adjoint versus exact adjoint operators

We now investigate the effects of using the pseudo and exact adjoint simulation operators described in Section VI. The validity of the exact adjoint property was verified multiple times in a double-precision environment through the dot-product test using random numbers as input. We first employ these two simulation methods to generate pre-stack RTM images. For a fair comparison between the results, the weighting operator $\mathbf{W}_m^{-1}$ was applied to both images. Figure 6 shows that pre-stack RTM, utilizing either simulation method, leads to comparable migration outcomes. The exact adjoint operator seems to provide slightly greater amplitudes in the deeper regions but the implementation of the exact adjoint simulation is significantly more difficult than the standard FDTD simulation. For the LSRTM problem, however, there is a significant difference between the results obtained using the pseudo and exact adjoint operators. Figure 7 presents data misfits of LSRTM with the different adjoints and weighting matrices. The solid blue and red lines in Fig. 7 show that employing the pseudo adjoint operator results in a much slower convergence compared to the exact adjoint operator. The dashed blue and red lines indicate that the regularization by using the operator $\mathbf{W}_m$ can accelerate convergence for both cases. Nevertheless, the pseudo adjoint operator still exhibits a slower convergence rate than the exact adjoint operator.

Figure 8 shows the effect of using the two types of adjoints in the image domain. In Fig. 8a we present the LSRTM image obtained using the pseudo adjoint operator, whereas Fig. 8b additionally incorporates the weighting operator $\mathbf{W}_m$. The two images are similar. Even though their resolution can exceed the resolution of the pre-stack RTM image, the complex structure enclosed by the red box is still insufficiently resolved for geological interpretation. In contrast, Figs. 8c and 8d show that the resolution of LSRTM can be considerably improved by adopting the exact adjoint operator.

Interestingly, we discovered that LSRTM, using both the pseudo adjoint operator and the non-identity weighting operator, exhibits a convergence curve remarkably similar to LSRTM with the exact adjoint operator and the identity weighting operator. The data misfit of the latter is only marginally lower. Initially, we anticipated that the LSRTM

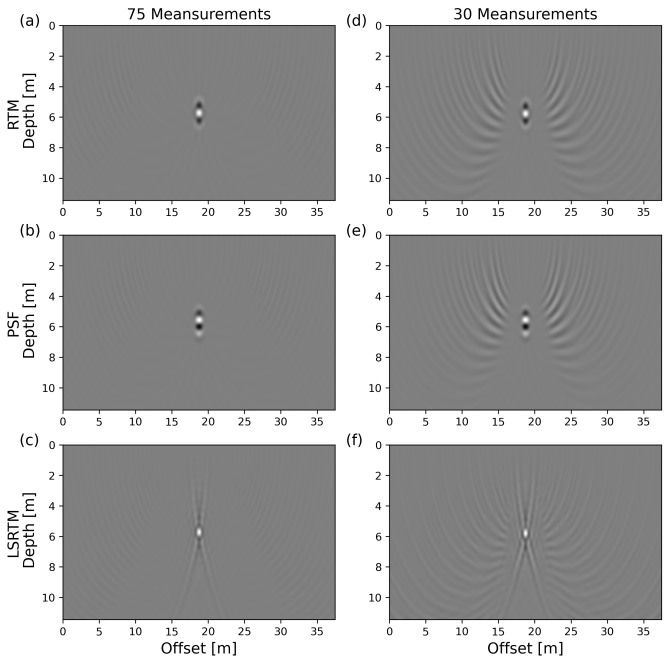

Fig. 10. Point scatterer test with 75 measurements along the surface (left-hand column) and 30 measurements along the surface (right-hand column): (a) and (d) pre-stack RTM, (b) and (e) LSRTM, and (c) and (f) point-spread function.

images produced by both methods could offer a comparable structural resolution. However, in contrast to the image produced with the pseudo adjoint operator and regularization, the LSRTM image generated with the exact adjoint operator demonstrates noticeably higher resolution in complex regions, as illustrated by the comparison of Figs. 8b and 8c. We suspect that this difference in image resolution is caused by the lower sensitivity to fine-scale structures. Large perturbations of low-sensitivity parameters can contribute only minimal changes in data. Conversely, even a slightly better data misfit can lead to much-improved resolution. Hence, this highlights the importance of reducing data misfits through the use of the exact adjoint operator in LSRTM.

Meanwhile, it is crucial to emphasize that data overfitting must be avoided with the LSRTM method. As discussed in section IV, the damping parameter is controlled by the number of iterations. An excessive amount of iterations can lead to poorly constrained artifacts within the null space. As depicted in Fig. 9, progressively increasing the number of iterations can introduce a substantial amount of artifacts. The artifacts are intrinsically caused by the limitations of the experimental design, meaning that they are the results of poorly resolved model parameters. Further details on the causes of these artifacts are explored in Section VIII-B.

B. Artifacts due to limited acquisition aperture

Figure 9d displays an 'X-shaped' artifact near the edges of the image. Intuitively, one might assume that this artifact is associated with the PML boundary condition. However, even after a comprehensive parameter tuning with regard to the PML, the artifact persists in the image.

To investigate the origin of this type of artifact, we conducted an additional synthetic test. In this test, a point scatterer was centrally located within the model. The background's relative permittivity was set to one and its conductivity and and the magnetic permeability were fixed at 10^{-3} S/m and 1 H/m. The point scatterer shared the same electric properties as the background, except for an increased permittivity of 20. The model was discretized on a 2D grid with a spatial sampling interval of 0.05 m. The point scatterer's dimensions were 0.05×0.05 m, equivalent to a single pixel in the model grid. We employed 75 zero-offset GPR measurements along the Earth's surface, each even spaced at 0.44 m apart, to illuminate the scatter. Our tests indicate that utilizing more than 75 sources does not considerably enhance the migration image quality.

Figure 10a presents the pre-stack RTM image of the point scatterer. While the image successfully depicts the scatterer's location, its precise position is blurred by side-lobes. To study blurring effect, we compute the point-spread function (PSF) of the point scatterer [23]. In our synthetic test, the exact location of the scatterer is known. This allows us to define a LSRTM model $\mathbf{m}$ as a zero matrix, except at the scatterer's exact location, where the value is set to one. Subsequently, the PSF of the scatterer can be derived by calculating $\mathbf{L}^T\mathbf{L}\mathbf{m}$. This process essentially is equivalent to extracting a specific row from the Hessian matrix $\mathbf{H} = \mathbf{L}^T\mathbf{L}$ that corresponds to the scatterer's location. Fig. 10b depicts the PSF. The high similarly between Figs. 10a and 10b indicates that the Hessian matrix can predict the blurring effect, which in turn can be expected to be removed by applying the inverse Hessian, that is, by using the LSRTM method.

Figure 10c shows the LSRTM image of the point scatterer. The side-lobe blurring has been successfully eliminated and thus the image has higher resolution and clearly depicts the location of the scatterer. Despite these improvements, the image is still contaminated by the aforementioned 'X-shaped' artifact, which can also be observed in least-squares Kirchhoff migration. Hu's research [72] explains that the shape of this artifact is associated with the acquisition geometry and the location of the scatterer. We additionally attribute this artifact to ambiguities within the model space. Locations of artifacts exhibit a significant trade-off with the actual scatterer location. This type of artifact cannot be entirely eliminated unless more information, such as a wider acquisition aperture or sparsity assumptions, that is, more prior information, are incorporated into the inverse problem. It is crucial to be aware of this type of artifact when targeting to image small-scale features.

Due to the absence of point scatterers, the 'X-shaped' artifact is not so obvious in our LSRTM image of the Marmousi model. Nevertheless, we believe that the 'X-shaped' artifact is related to the artifacts appearing near the left- and right-hand edges of the survey area in Fig. 9d. Layers are cut off abruptly at the boundaries and thus causing the reflectors at the boundaries to behave like scatterers. At the boundaries, the acquisition aperture is limited. Therefore, the 'x-shape' artifacts appear particularly prominently.

C. Artifacts due to coarse acquisition spacing

A coarse acquisition spacing can pose a significant concern in 3D surveys. Despite a small spacing between the measurement points in the inline direction, the trace density in the cross-line direction can be significantly sparser. To study the effects of coarse acquisition, we conducted an additional point-scatterer test, similar to the previous one, but with the number of sources reduced to 30, resulting in a spatial sampling interval of 1.1 m. The RTM image, PSF, and LSRTM image resulting from this test are displayed in Figs. 10 (d)-(f). Coarse acquisition results in artifacts near the point scatterer, resembling the Gibbs phenomenon. Unfortunately, these artifacts cannot be suppressed by LSRTM. The artifacts are prominent in the near-surface region but diminish with depth. We believe this is due to the fact that the point-sources can merge their wavefronts to form plane waves at greater depths. This could possibly explain the near-surface artifacts along the first layer in Fig. 9d.

In the synthetic test, the artifact can be effortlessly eliminated using a damping filter, given that the first layer is flat. For models with more complex structures, we anticipate that a data interpolation algorithm is needed to mitigate these artifacts.

D. Computational cost

With regard to the computational cost, LSRTM is undoubtedly more demanding than RTM. Each iteration of LSRTM requires the execution of both forward and adjoint operations, which entails two times the wave simulation per source. To save computing time, we reduce the number of iterations by using the aforementioned source wavefield illumination-compensation regularization [65] in addition to using a GPU cluster and parallelizing the simulations for all sources.

We utilized 28 computing nodes, each equipped with 4 NVIDIA GeForce GTX TITAN X GPU devices, to implement the simulations. On average, the forward and adjoint operations took approximately 155 and 230 seconds for 200 sources, respectively. We do not purport that our code is fully optimized, but we underscore that, while LSRTM has a higher computational demand than RTM approaches, modern computational resources have made 2D LSRTM quite affordable, which, in turn, opens the perspective of extending the presented approach to 3D.

IX. Conclusion

This study develops and explores pre-stack LSRTM for GPR imaging, with the goal of enhancing image resolution. The forward operator of LSRTM is derived using the Born approximation of Maxwell's equations, and its exact adjoint operator is derived via matrix operations. Employing these two operators, we obtain the LSRTM image using the conjugate gradients method. Synthetic examples demonstrate that LSRTM provides superior resolution compared to images generated by both pre-stack and post-stack RTM methods, thus, improving the visibility of fine-scale structures. We conduct a point-scatterer test to evidence that the improved resolution can be attributed to the elimination of the point-spreading function.

We put particular emphasis on the derivation of the adjoint operator, ensuring its ability to pass the dot-product test with machine precision. Fulfilling this criterion necessitates the use of an exact adjoint simulation that is not the same as the pseudo adjoint used for the standard time-reversal simulation. To establish the correct formulation, it was necessary to express the FDTD simulation with CPML boundary conditions in matrix form and subsequently apply a conjugate transpose to the entire equation. Our synthetic tests demonstrate that the exact adjoint operator led to significantly faster convergence and to a lower data misfit.

Two issues can negatively affect LSRTM image quality: limited acquisition aperture and coarse acquisition spacing. The former can cause 'X-shaped' artifacts in the images and the latter can create artifacts resembling the Gibbs phenomenon. Unfortunately, both types of artifacts cannot be suppressed by the least-squares inversion because they are affected by the poorly resolved model parameters, that is, the so-called null-space. We expect that the artifacts can be intrinsically reduced by incorporating more collected or reconstructed data into the inverse problem.

Finally, although LSRTM has undoubtedly a higher computational cost compared to RTM, we have shown that modern computational devices render 2D LSRTM quite affordable, thus offering the perspective on an eventual extension to 3D. Implementing 3D LSRTM can pose significant challenges, primarily due to its immense computational demands that can exceed the memory capacity of GPUs. Therefore, effective domain partitioning for inversion and efficient reconstruction of source wavefields will be crucial research areas. Additionally, resolving ill-posedness caused by the poor sampling in the cross-line direction and optimizing data acquisition design for 3D surveys are highly anticipated research areas.

Appendix A
Finite-difference Stepping Matrix with CPML Boundary Condition

The matrix $\mathbf{T}$ in Eq. (56) can be expressed by the multiplication of the following four matrices

$$\mathbf{T}_4 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{C}_{bz}\mathbf{F}_{zb} & \mathbf{C}_{bx}\mathbf{F}_{xb} & \mathbf{C}_a & \mathbf{0} & \mathbf{0} & \mathbf{C}_c & -\mathbf{C}_c \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (67)$$

$$\mathbf{T}_3 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_x\mathbf{F}_{xb} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_x & \mathbf{0} \\ \mathbf{A}_z\mathbf{F}_{zb} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_z \end{bmatrix}, \quad (68)$$

$$\mathbf{T}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & -\mathbf{D}_{bz}\mathbf{F}_{zf} & -\mathbf{D}_c & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{D}_{bx}\mathbf{F}_{xf} & \mathbf{0} & \mathbf{D}_c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{69}$$

and

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_z\mathbf{F}_{zf} & \mathbf{B}_z & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_x\mathbf{F}_{xf} & \mathbf{0} & \mathbf{B}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{70}$$

After multipling the four matrices $\mathbf{T} = \mathbf{T}_4\mathbf{T}_3\mathbf{T}_2\mathbf{T}_1$, we have a compact form of $\mathbf{T}$ that is not a self-adjoint matrix.

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & -\mathbf{A}_z\mathbf{F}_{zf}\mathbf{D}_c - \mathbf{D}_{bz}\mathbf{F}_{zf} \\ \mathbf{0} & \mathbf{I} & \mathbf{A}_x\mathbf{F}_{xf}\mathbf{D}_c + \mathbf{D}_{bx}\mathbf{F}_{xf} \\ -\mathbf{C}_{bz}\mathbf{F}_{zb} & \mathbf{C}_{bx}\mathbf{F}_{xb} & \begin{array}{c}\mathbf{A}_z\mathbf{F}_{zf}\mathbf{D}_c\mathbf{C}_{bz}\mathbf{F}_{zb}+\mathbf{A}_x\mathbf{F}_{xf}\mathbf{D}_c\mathbf{C}_{bx}\mathbf{F}_{xb} \\ +\mathbf{D}_{bx}\mathbf{F}_{xf}\mathbf{C}_{bx}\mathbf{F}_{xb} \\ +\mathbf{D}_{bz}\mathbf{F}_{zf}\mathbf{C}_{bz}\mathbf{F}_{zb}+\mathbf{C}_a\end{array} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_z\mathbf{F}_{zf} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_x\mathbf{F}_{xf} \\ \mathbf{0} & \mathbf{A}_x\mathbf{F}_{xb} & \mathbf{A}_x\mathbf{F}_{xf}\mathbf{D}_c\mathbf{A}_x\mathbf{F}_{xb} + \mathbf{D}_{bx}\mathbf{F}_{xf}\mathbf{A}_x\mathbf{F}_{xb} \\ \mathbf{A}_z\mathbf{F}_{zb} & \mathbf{0} & -\mathbf{A}_z\mathbf{F}_{zf}\mathbf{D}_c\mathbf{A}_z\mathbf{F}_{zb} - \mathbf{D}_{bz}\mathbf{F}_{zf}\mathbf{A}_z\mathbf{F}_{zb} \end{bmatrix}$$

$$\begin{bmatrix} -\mathbf{B}_z\mathbf{D}_c & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_x\mathbf{D}_c & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_z\mathbf{D}_c\mathbf{C}_{bz}\mathbf{F}_{zb} & \mathbf{B}_x\mathbf{D}_c\mathbf{C}_{bx}\mathbf{F}_{xb} & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_z & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_x\mathbf{D}_c\mathbf{A}_x\mathbf{F}_{xb} & \mathbf{B}_x & \mathbf{0} \\ -\mathbf{B}_z\mathbf{D}_c\mathbf{A}_z\mathbf{F}_{zb} & \mathbf{0} & \mathbf{0} & \mathbf{B}_z \end{bmatrix} \tag{71}$$

ACKNOWLEDGMENT

L.X. gratefully acknowledges the funding provided by the Faculty of Geoscience and Environment at the University of Lausanne through the Junior Research Fellowship program.

REFERENCES

[1] R. Knight, "Ground penetrating radar for environmental applications," *Annual Review of Earth and Planetary Sciences*, vol. 29, no. 1, pp. 229–255, 2001.

[2] G. P. Radar, "Ground penetrating radar: theory and applications," *Jol ISBN*, pp. 978–0, 2009.

[3] A. Benedetto and L. Pajewski, *Civil engineering applications of ground penetrating radar*. Springer, 2015.

[4] A. P. Annan, "Gpr methods for hydrogeological studies," in *Hydrogeophysics*. Springer, 2005, pp. 185–213.

[5] L. A. Plewes and B. Hubbard, "A review of the use of radio-echo sounding in glaciology," *Progress in physical geography*, vol. 25, no. 2, pp. 203–236, 2001.

[6] F. Navarro and O. Eisen, "11 ground-penetrating radar in glaciological applications," in *Remote sensing of glaciers: Techniques for topographic, spatial and thematic mapping of glaciers*. Taylor & Francis London, 2009, pp. 195–229.

[7] L. B. Conyers, *Ground-penetrating radar for archaeology*. Rowman & Littlefield, 2023.

[8] A. Neal, "Ground-penetrating radar and its use in sedimentology: principles, problems and progress," *Earth-science reviews*, vol. 66, no. 3-4, pp. 261–330, 2004.

[9] S. H. Gray, "Efficient traveltime calculations for Kirchhoff migration," *Geophysics*, vol. 51, no. 8, pp. 1685–1688, 1986. [Online]. Available: http://dx.doi.org/10.1190/1.1442217

[10] R. H. Stolt, "Migration by fourier transform," *Geophysics*, vol. 43, no. 1, pp. 23–48, 1978. [Online]. Available: https://doi.org/10.1190/1.1440826

[11] J. Zhu and L. R. Lines, "Comparison of kirchhoff and reverse-time migration methods with applications to prestack depth imaging of complex structures," *Geophysics*, vol. 63, no. 4, pp. 1166–1176, 1998. [Online]. Available: https://doi.org/10.1190/1.1444416

[12] J. Etgen, S. H. Gray, and Y. Zhang, "An overview of depth imaging in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCA5–WCA17, 2009.

[13] A. Giannopoulos, "Modelling ground penetrating radar by gprmax," *Construction and Building Materials*, vol. 19, no. 10, pp. 755–762, 2005, non Destructive Testing: Selected papers from Structural Faults and Repair 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950061805001509

[14] J. Irving and R. Knight, "Numerical modeling of ground-penetrating radar in 2-d using matlab," *Computers & Geosciences*, vol. 32, no. 9, pp. 1247–1258, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098300405002621

[15] D. Loewenthal, L. Lu, R. Roberson, and J. Sherwood, "The wave equation applied to migration," *Geophysical Prospecting*, vol. 24, no. 2, pp. 380–399, 1976. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2478.1976.tb00934.x

[16] G. McMechan, "Migration by extrapolation of time-dependent boundary values," *Geophysical Prospecting*, vol. 31, no. 3, pp. 413–420, 1983.

[17] E. Fisher, G. A. McMechan, A. P. Annan, and S. W. Cosway, "Examples of reverse-time migration of single-channel, ground-penetrating radar profiles," *Geophysics*, vol. 57, no. 4, pp. 577–586, 1992. [Online]. Available: https://doi.org/10.1190/1.1443271

[18] T. Zhu, J. M. Carcione, and M. A. Botelho, "Reverse time imaging of ground-penetrating radar and sh-seismic data including the effects of wave loss," *Geophysics*, vol. 81, no. 4, pp. H21–H32, 2016.

[19] Q. Di and M. Wang, "Migration of ground-penetrating radar data with a finite-element method that considers attenuation and dispersion," *Geophysics*, vol. 69, no. 2, pp. 472–477, 2004.

[20] X. Wang, T. Yu, D. Feng, S. Ding, B. Li, Y. Liu, and Z. Feng, "A high-efficiency spectral element method based on cfs-pml for gpr numerical simulation and reverse time migration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 1232–1243, 2023.

[21] J. H. Bradford, J. Privette, D. Wilkins, and R. Ford, "Reverse-time migration from rugged topography to image ground-penetrating radar data in complex environments," *Engineering*, vol. 4, no. 5, pp. 661–666, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2095809918304594

[22] Y. Zhang and J. Sun, "Practical issues in reverse time migration: True amplitude gathers, noise removal and harmonic source encoding," *First break*, vol. 27, no. 1, 2009.

[23] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory*. San Diego, CA: Academic Press, 1989.

[24] P. Sava and I. Vasconcelos, "Extended imaging conditions for wave-equation migration," *Geophysical Prospecting*, vol. 59, no. 1, pp. 35–55, 2011.

[25] J. F. Claerbout, "Toward a unified theory of reflector mapping," *Geophysics*, vol. 36, no. 3, pp. 467–481, 1971.

[26] J. Claerbout, *Earth Soundings Analysis: Processing Versus Inversion*, ser. Blackwell scientific publications. Blackwell Scientific Publications, 1992. [Online]. Available: https://books.google.ch/books?id=ws1qQgAACAAJ

[27] D. Leparoux, D. Gibert, and P. Côte, "Adaptation of prestack migration to multi-offset ground-penetrating radar (gpr) data," *Geophysical prospecting*, vol. 49, no. 3, pp. 374–386, 2001.

[28] F. Liu, G. Zhang, S. A. Morton, and J. P. Leveille, "An effective imaging condition for reverse-time migration using wavefield decomposition," *Geophysics*, vol. 76, no. 1, pp. S29–S39, 2011. [Online]. Available: http://dx.doi.org/10.1190/1.3533914

[29] B. D. Nguyen and G. A. McMechan, "Excitation amplitude imaging condition for prestack reverse-time migration," *Geophysics*, vol. 78, no. 1, pp. S37–S46, 2013. [Online]. Available: https://doi.org/10.1190/geo2012-0079.1

[30] J. Huo, Q. Zhao, B. Zhou, L. Liu, C. Ma, J. Guo, and L. Xie, "Energy flow domain reverse-time migration for borehole radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7221–7231, 2019.

[31] D. Feng, Z. Feng, X. Wang, D. Xu, Y. Liu, B. Li, T. Yu, S. Ding, and T. Li, "Reverse time migration of ground penetrating radar with optimized full wavefield separation based on poynting vector imaging condition and tv-l1 based artifacts suppression," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[32] C. Zeng, S. Dong, J. Mao, and B. Wang, "Broadband least-squares reverse time migration for complex structure imaging," in *SEG International Exposition and Annual Meeting*. SEG, 2014, pp. SEG–2014.

[33] C. Zeng, S. Dong, and B. Wang, "Least-squares reverse time migration: Inversion-based imaging toward true reflectivity," *The Leading Edge*, vol. 33, no. 9, pp. 962–968, 2014.

[34] Z. Xue, Y. Chen, S. Fomel, and J. Sun, "Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization," *Geophysics*, vol. 81, no. 1, pp. S11–S20, 2016.

[35] R.-E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 11 2006. [Online]. Available: https://doi.org/10.1111/j.1365-246X.2006.02978.x

[36] R. Keys and A. Weglein, "Generalized linear inversion and the first Born theory for acoustic media," *Journal of Mathematical Physics*, vol. 24, no. 6, pp. 1444–1449, 1983.

[37] P. Lailly, "The seismic inverse problem as a sequence of before stack migrations," *Society of Industrial and Applied Mathematics, Expanded Abstracts*, pp. 206–220, 1983.

[38] W. B. Beydoun and M. Mendes, "Elastic ray-born l 2-migration/inversion," *Geophysical Journal International*, vol. 97, no. 1, pp. 151–160, 1989.

[39] T. Nemeth, C. Wu, and G. T. Schuster, "Least-squares migration of incomplete reflection data," *Geophysics*, vol. 64, no. 1, pp. 208–221, 1999. [Online]. Available: http://dx.doi.org/10.1190/1.1444517

[40] B. Duquet, K. J. Marfurt, and J. A. Dellinger, "Kirchhoff modeling, inversion for reflectivity, and subsurface illumination," *Geophysics*, vol. 65, no. 4, pp. 1195–1209, 2000.

[41] H. Kühl and M. D. Sacchi, "Least-squares wave-equation migration for AVP/AVA inversion," *Geophysics*, vol. 68, no. 1, pp. 262–273, 2003. [Online]. Available: http://dx.doi.org/10.1190/1.1543212

[42] M. L. Clapp, "Imaging under salt: Illumination compensation by regularized inversion," Ph.D. dissertation, Stanford University, 2005.

[43] J. Wang, H. Kuehl, and M. D. Sacchi, "High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the Western Canadian Sedimentary Basin," *Geophysics*, vol. 70, no. 5, pp. S91–S99, 2005. [Online]. Available: http://dx.doi.org/10.1190/1.2076748

[44] *Structure Constrained Least-squares Migration*, ser. SEG International Exposition and Annual Meeting, vol. All Days, 10 2009.

[45] W. Dai, X. Wang, and G. T. Schuster, "Least-squares migration of multisource data with a deblurring filter," *Geophysics*, vol. 76, no. 5, pp. R135–R146, 2011. [Online]. Available: http://dx.doi.org/10.1190/geo2010-0159.1

[46] Z. Xue, Y. Chen, S. Fomel, and J. Sun, "Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization," *Geophysics*, vol. 81, no. 1, pp. S11–S20, 2016. [Online]. Available: https://doi.org/10.1190/geo2014-0524.1

[47] L. Xu and M. D. Sacchi, "Preconditioned acoustic least-squares two-way wave-equation migration with exact adjoint operator," *Geophysics*, vol. 83, no. 1, pp. S1–S13, 2018. [Online]. Available: https://doi.org/10.1190/geo2017-0167.1

[48] P. Wang, S. Huang, and M. Wang, "Improved subsalt images with least-squares reverse time migration," *Interpretation*, vol. 5, no. 3, pp. SN25–SN32, 2017.

[49] J. Carcione and F. Cavallini, "On the acoustic-electromagnetic analogy," *Wave motion*, vol. 21, no. 2, pp. 149–162, 1995.

[50] L. Xu and M. Sacchi, "Least squares reverse time migration with model space preconditioning and exact forward/adjoint pairs," *78th EAGE Conference and Exhibition*, 2016.

[51] W. C. Chew and W. H. Weedon, "A 3d perfectly matched medium from modified maxwell's equations with stretched coordinates," *Microwave and optical technology letters*, vol. 7, no. 13, pp. 599–604, 1994.

[52] D. Komatitsch and R. Martin, "An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation," *Geophysics*, vol. 72, no. 5, pp. SM155–SM167, 2007.

[53] R. Versteeg, "The marmousi experience: Velocity model determination on a synthetic complex data set," *The Leading Edge*, vol. 13, no. 9, pp. 927–936, 1994. [Online]. Available: https://doi.org/10.1190/1.1437051

[54] A. Tarantola, "Generalized nonlinear inverse problems solved using the least squares criterion," *Reviews of Geophysics*, vol. 20, no. 2, pp. 219–232, 1982. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG020i002p00219

[55] J. F. Claerbout and I. Green, "Basic earth imaging," 2008.

[56] J. Schleicher, P. Hubral, M. Tygel, and M. S. Jaya, "Minimum apertures and fresnel zones in migration and demigration," *Geophysics*, vol. 62, no. 1, pp. 183–194, 1997.

[57] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Society for Industrial Mathematics, 1987.

[58] A. Guitton, A. Valenciano, and D. Bevc, "Robust imaging condition for shot-profile migration," *SEG Technical Program Expanded Abstracts*, pp. 2519–2523, 2006. [Online]. Available: http://library.seg.org/doi/abs/10.1190/1.2370043

[59] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, 1952.

[60] J. A. Scales, "Tomographic inversion via the conjugate gradient method," *Geophysics*, vol. 52, no. 2, pp. 179–185, 1987. [Online]. Available: https://doi.org/10.1190/1.1442293

[61] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on antennas and propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[62] B. Fornberg, "Generation of finite difference formulas on arbitrarily spaced grids," *Mathematics of computation*, vol. 51, no. 184, pp. 699–706, 1988.

[63] R. J. Luebbers and F. Hunsberger, "Fdtd for nth-order dispersive media," *IEEE transactions on Antennas and Propagation*, vol. 40, no. 11, pp. 1297–1301, 1992.

[64] G. Chavent, *Nonlinear least squares for inverse problems*. Springer, 2009.

[65] J. E. Rickett, "Illumination-based normalization for wave-equation depth migration," *Geophysics*, vol. 68, no. 4, pp. 1371–1379, 2003. [Online]. Available: https://doi.org/10.1190/1.1598130

[66] Y. Liu, J. Irving, and K. Holliger, "High-resolution velocity estimation from surface-based common-offset GPR reflection data," *Geophysical Journal International*, vol. 230, no. 1, pp. 131–144, 02 2022. [Online]. Available: https://doi.org/10.1093/gji/ggac058

[67] X. Comas, L. Slater, and A. Reeve, "In situ monitoring of free-phase gas accumulation and release in peatlands using ground penetrating radar (gpr)," *Geophysical research letters*, vol. 34, no. 6, 2007.

[68] R. W. Jacob and T. Urban, "Ground-penetrating radar velocity determination and precision estimates using common-midpoint (cmp) collection with hand-picking, semblance analysis and cross-correlation analysis: A case study and tutorial for archaeologists," *Archaeometry*, vol. 58, no. 6, pp. 987–1002, 2016.

[69] H. Liu, Z. Yang, Y. Yue, X. Meng, C. Liu, and J. Cui, "Asphalt pavement characterization by gpr using an air-coupled antenna array," *NDT & E International*, vol. 133, p. 102726, 2023.

[70] L. Xu, E. Manukyan, and H. Maurer, "Optimized model parametrization using compact full waveform inversion," *Geophysical Journal International*, vol. 222, no. 1, pp. 628–645, 04 2020. [Online]. Available: https://doi.org/10.1093/gji/ggaa175

[71] L. Xu, V. Winner, and H. Maurer, "Gradient-constrained model parametrization in 3-d compact full waveform inversion," *Geophysical Journal International*, vol. 232, no. 1, pp. 366–397, 2023.

[72] J. Hu, G. T. Schuster, and P. A. Valasek, "Poststack migration deconvolution," *Geophysics*, vol. 66, no. 3, pp. 939–952, 2001.

Linan Xu serves as a Junior Lecturer at the University of Lausanne. Xu earned his Bachelor's and Master's degrees from the University of Alberta and completed his Ph.D. at ETH Zürich. His research focuses on advancing detection capabilities in reflection seismology and radar imaging. His research interests include PDE-constrained nonlinear optimization, finite-difference modelling, and 3D full waveform inversion and migration techniques.

James Irving received the bachelor's degree in Earth sciences from the University of Waterloo, Waterloo, ON, Canada, in 1997, the master's degree in geophysics from the University of British Columbia, Vancouver, BC, Canada, in 2000, and the Ph.D. degree in geophysics from Stanford University, Stanford, CA, USA, in 2006. He is currently an Associate Professor of applied and environmental geophysics with the University of Lausanne, Lausanne, Switzerland.

Klaus Holliger received the M.Sc. and Ph.D. degrees in geophysics from ETH Zürich, Switzerland, in 1987 and 1991, respectively, and the master's degree in economics from the University of London, U.K., in 2000. He held a post-doctoral position with Rice University, Houston, TX, USA, before joining ETH Zurich as a Lecturer. He currently holds a professorship with the University of Lausanne, Switzerland. His primary research interests include hydrogeophysics sensu lato.

Dr. Holliger has received the Harold Money Award and Frank Frischknecht Leadership Award of the Society of Exploration Geophysicists (SEG). As a founding member of the Applied and Environmental Geophysics Group, ETH Zurich, he also shared the SEG's Distinguished Service Award. He has served as a President of the Near Surface Geophysics Section, SEG, an Editor-in-Chief of the Journal of Applied Geophysics, an Associate Editor of Geophysics, Water Resources Research, Geology, and Solid Earth, and a Co-Editor of several books and special issues. Most recently, he was the SEG Distinguished Lecturer Europe.