FULL WAVEFORM INVERSION AND LAGRANGE MULTIPLIERS

Ali Gholami
Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland
agholami@igf.edu.pl

November 21, 2023

ABSTRACT

Full-waveform inversion (FWI) is an effective method for imaging subsurface properties using sparsely recorded data. It involves solving a wave propagation problem to estimate model parameters that accurately reproduce the data. Recent trends in FWI have led to the development of extended methodologies, among which source extension methods leveraging reconstructed wavefields to solve penalty or augmented Lagrangian (AL) formulations have emerged as robust algorithms, even for inaccurate initial models. Despite their demonstrated robustness, challenges remain, such as the lack of a clear physical interpretation, difficulty in comparison, and reliance on difficult-to-compute least squares (LS) wavefields. This paper is divided into two critical parts. In the first, a novel formulation of these methods is explored within a unified Lagrangian framework. This novel perspective permits the introduction of alternative algorithms that employ LS multipliers instead of wavefields. These multiplier-oriented variants appear as regularizations of the standard FWI, are adaptable to the time domain, offer tangible physical interpretations, and foster enhanced convergence efficiency. The second part of the paper delves into understanding the underlying mechanisms of these techniques. This is achieved by solving the FWI equations using iterative linearization and inverse scattering methods. The paper provides insight into the role and significance of Lagrange multipliers in enhancing the linearization of FWI equations. It explains how different methods estimate multipliers or make approximations to increase computing efficiency. Additionally, it presents a new physical understanding of the Lagrange multiplier used in the AL method, highlighting how important it is for improving algorithm performance when compared to penalty methods.

Keywords: Extended Full Waveform inversion; Search space extension; Physical meaning of Lagrange multipliers; Wavefield reconstruction inversion; Multiplier waveform inversion.

1 Introduction

Full waveform inversion (FWI) is a widely employed and effective technique for imaging subsurface properties in geophysical studies [22, 24, 29, 37]. This involves solving a wave propagation problem to determine the model parameters that accurately reproduce the recorded/observed data. The governing law is defined by a set of nonlinear equations that capture the wavefields and establish their connection to the model parameters, thereby enabling the calculation of synthetic data. The primary objective of FWI is to solve this set of nonlinear equations to estimate the model parameters based on observed data.

Unlike linear systems, such nonlinear systems of equations cannot be solved by direct methods due to their complexity. Therefore, they must be solved using iterative methods. These iterative methods can be generated based on various principles, such as iterative linearization approaches or inverse problem theory. In the iterative linearization approach, the nonlinear system is linearized at each iteration, and the linearized equations are then solved to obtain an updated solution. This process is repeated until convergence is achieved. On the other hand, the inverse problem theory approach replaces the problem of solving the system of nonlinear equations with the optimization of some nonlinear functional. The desired solution is then obtained by finding the best solution in a precise sense, usually by minimizing

the difference between the observed data and the predicted data [32]. This paper is divided into two independent sections, each of which focuses on one of these frameworks of solving the FWI equations.

In a first section, we consider the traditional form of FWI in the framework of inverse problem theory. It addresses the estimation of subsurface model parameters by minimizing the difference between observed and synthetic data through a nonlinearly constrained least-squares (LS) optimization problem. To incorporate physical constraints and prior information for regularization, the widely adopted Lagrange multiplier method is employed [13]. This approach introduces Lagrange multipliers to enforce constraints such as the wave equation governing the wave propagation problem. The desired solution, which encompasses the model parameters, wavefields, and multipliers, is obtained by solving three sets of nonlinear equations that satisfy the Karush-Kuhn-Tucker (KKT) optimality conditions [21]. These conditions are satisfied by setting the partial derivatives of the Lagrangian with respect to the multipliers, wavefields, and model parameters equal to zero.

The solution of the KKT equations can be achieved through Newton-type methods. These methods iteratively linearize the system around the current solution estimate and solve the resulting linear system using preconditioned Krylov iterative methods [2, 14]. However, efficiently solving large KKT systems poses significant challenges from both computational and memory perspectives. The large size of the systems can lead to high computational costs and memory requirements, especially for large-scale problems encountered in practical applications. Furthermore, the Hessian matrix associated with the KKT system may be ill-conditioned and indefinite, leading to numerical instabilities and convergence issues. To address the ill-conditioning, appropriate preconditioning or regularization procedures are needed, as well as effective ways for memory issues.

The memory issue can be mitigated through variable projection [11]. Traditionally, this approach involves eliminating the wavefields and multipliers from the optimization variables by exactly satisfying two sets of the KKT equations, the forward and adjoint wave equations. The gradient of the reduced objective function, necessary for updating the model parameters, is then computed by correlating the wavefields and multipliers, which propagate forward and backward in time, respectively [23]. While this reduced approach offers significant computational and memory savings, it still faces challenges related to the ill-conditioning of the inversion process, leading to numerical instabilities and slow convergence [19]. Recent advancements in computational resources have enabled solving large and complex FWI problems using advanced numerical optimization methods. As a result, addressing the ill-conditioning issue and improving the convergence behavior of FWI methods have become an active areas of research [e.g., 1, 3, 4, 7, 15, 16, 20, 22, 25, 36, 39, 41].

This study focuses on investigating FWI methods within a Lagrangian framework, including the penalty and augmented Lagrangian (AL) methods. In its standard form, the penalty method enforces constraints by adding a quadratic term to the objective function that penalizes the sum of the squares of the wave-equation violations. The amount of penalization is determined by the penalty parameter, which must be increased with iteration to satisfy the constraints. To solve the penalty objective function, van Leeuwen & Herrmann [36] employed the Wavefield Reconstruction Inversion (WRI) algorithm, which follows a two-step alternating iteration for updating the wavefields and model parameters. The main issue with WRI is that it requires high values of the penalty parameter to satisfy the wave equation, yet high values slow down the algorithm's convergence because this method essentially employs a step length that varies with the inverse of the penalty parameter. To solve this issue, the AL method combines both the Lagrangian and penalty methods. This results in the iteratively refined WRI (IR-WRI) algorithm, similar to WRI but with a modified volume source term [1]. The AL method satisfies the wave equation even for moderate values of the penalty parameter without the need for excessively large values. This preserves the well-posedness of the algorithm through iterations which leads to accurate estimations of the multiplier and to an improved convergence behavior.

In recent years, numerous papers have demonstrated the superior performance of Lagrangian-based methods, such as WRI and IR-WRI, compared to standard FWI, using various numerical examples [e.g. 1, 3, 7, 15, 17, 22, 25, 36]. However, despite their effectiveness, these methods face several challenges: (1) They are based on least squares (LS) wavefields, also known as extended wavefields or data-assimilated wavefields. These wavefields are obtained by solving an augmented wave equation that simultaneously incorporates the wave equation and the data equation in the LS sense. Efficient computation of these LS wavefields in time-domain applications is very challenging because explicit time-stepping methods for solving the augmented wave equation are lacking [see 7, 17, 22, 25, 38, and references therein] (2) Unlike traditional FWI methods, these approaches do not involve a backward wavefield. However, the model update still relies on correlating the LS wavefields and a quantity that is a function of these wavefields whose physical meaning is not immediately clear. (3) Interpreting the resulting algorithms physically can be challenging, making the mechanism of each method unclear or sometimes misunderstood. This lack of clarity makes it difficult to compare WRI and IR-WRI with each other and with standard FWI. This issue is mainly because they have not been considered under a unified formulation, and the physical significance of Lagrange multipliers and their function in these methods are not well understood.

To address these challenges, in this study, by investigating these methods within a unified framework, we aim to shed light on the underlying mechanisms and explore their potential for more practical applications. Understanding the physical meaning of Lagrange multipliers and their role in the inversion process not only enhances our theoretical understanding of the methods but also opens up new avenues for improving their performance and applicability in practice. We demonstrate that there exist alternative formulation for WRI and IR-WRI which gives the same solution to the inverse problem, but they are based on the computation of LS multipliers rather than LS wavefields. They lead to inversion algorithms that are extremely similar to the standard FWI, can be performed in the time domain using explicit time-stepping methods, and are physically interpretable. To achieve this, we formulate the FWI problem as a two-step iterative process. The first step involves solving a convex-concave min-max optimization problem to compute the wavefields and multipliers. Subsequently, the model parameters are updated in the second step using a correlation of the computed wavefields and multipliers, akin to the standard FWI. Solving the first step requires handling a two-by-two block linear system of equations, also known as a saddle point system. The penalty and AL methods are two different approaches for regularization of the ordinary Lagrangian hence enable the stable solution of this saddle point system. For solving this system, two distinct approaches can be employed: the wavefield-oriented approach and the multiplier-oriented approach. These approaches emphasize either wavefield reconstruction or multiplier computation as the primary goal, providing flexibility in implementing the algorithms.

The wavefield-oriented approach, which reduces to WRI in the penalty method [36] and to IR-WRI in the AL method [1], focuses on reconstructing the LS wavefields as an initial step. This approach aims to obtain accurate wavefield information, which is then used to derive the Lagrange multipliers. By utilizing the LS wavefields, multipliers can be estimated cheaply and used to calculate the model update. By contrast, the multiplier-oriented approach prioritizes the calculation of LS multipliers as the primary step. It utilizes these multipliers as secondary volume sources in the wave equation to calculate the wavefields. The advantage of this approach is that it allows the use of standard time-stepping methods to compute the wavefields and multipliers.

In the second part of this paper, we examine how to solve the original nonlinear equations that govern FWI without using optimization theory. Instead, we solved the FWI equations using scattering concepts and iterative linearization methods, aiming to find a set of model parameters and wavefields that satisfy the wave equation while remaining consistent with the recorded data. In this framework, Lagrange multipliers are introduced as auxiliary variables that split the nonlinear equations, hence simplifying the linearization process. This allows us to shed light on the physical interpretation of Lagrange multipliers, revealing their role as scattering sources or scattered data, which significantly contributes to the linearization of FWI equations and improves the overall iteration efficiency. We demonstrate how various FWI methods strive to estimate accurate Lagrange multipliers or employ approximations to enhance computational and memory efficiency. Importantly, this study offers a pioneering insight into the physical meaning of the Lagrange multiplier used in the AL method, emphasizing its significance in enhancing algorithm performance compared to penalty method.

2 NOTATION

In this paper, we develop iterative methods for solving FWI. We use a streamlined notation to represent the iterative process to maintain a clear and concise presentation throughout the analysis. In particular, we denote the background values of the variables without indices, while the updated variables are denoted by a superscript "+". In some circumstances, it may be necessary to construct updated variables using the same variables used to create background variables. The superscript "−" is used to identify these variables. For instance, we input the variables' background values (m and u), as well as their previous values (m^- and u^-) for each iteration. The updated values of the variables, denoted as m^+ and u^+, are obtained after performing the necessary computations and updates. The updated variables from the previous iteration serve as the background variables for the subsequent iteration, which adheres to the same convention. We hope to simplify the presentation and make the iterative process easier to understand by using this notation. It keeps the text and formulas simple, and allows us to concentrate on the key components of each iteration.

3 Lagrangian based full waveform inversion

We start with the assumption that the seismic wavefield, denoted by $u(t, x; x_s)$, is generated by a source term $b(t, x; x_s)$ in a medium characterized by variable velocity and constant density. This wavefield satisfies the wave equation given by

$$m(x)\frac{\partial^2}{\partial t^2}u(t, x; x_s) - \nabla^2 u(t, x; x_s) = b(t, x; x_s), \tag{1}$$

where $m(x)$ represents the model parameters, specifically, the squared slowness, which varies with the spatial coordinate x in two or three dimensions. Source locations are denoted by x_s for $s = 1, 2, ..., N_s$, where N_s is the total number of sources. Operator ∇^2 represents the Laplacian.

In practical applications, the wavefield is typically sampled at the receiver locations x_r for $r = 1, 2, ..., N_r$ (where N_r is the number of receivers). These receivers can be positioned in various ways, depending on the imaging objectives and characteristics of the subsurface under investigation. Examples include placing receivers in boreholes, deploying them on the seafloor, and distributing them throughout the survey area to capture the seismic wavefield as it propagates through the subsurface. The recorded samples at these receiver locations, denoted by $d(t, x_r; x_s)$, are equivalent to $u(t, x_r; x_s)$ and serve as the input data for the FWI algorithms to estimate the model parameters.

For simplicity, our analysis is based on single-source problems. However, this does not impose any limitations on our formulation because the extension to multiple sources can be achieved by summing the final model update formula over the sources.

Assuming that a known source b^* is used to generate the true wavefield u^* in a medium characterized by model parameters m^*, we can record the corresponding data d^*. These quantities are connected by a pair of coupled equations.

$$A(m^*)u^* = b^*, \tag{2a}$$
$$Pu^* = d^*. \tag{2b}$$

Here, $A(m) \in \mathbb{R}^{N \times N}$ represents the discretization of wave equation (1) with appropriate boundary conditions. The matrix size N is determined by $N_m \times N_t$, where N_m is the number of grid points used to sample the subsurface model and N_t is the number of time samples. Meanwhile, $P \in \mathbb{R}^{M \times N}, M = N_r \times N_t$, represents the sampling operator, which samples the wavefield at the receiver locations x_r.

To solve nonlinear equations (2a) and (2b), they are often transformed into a nonlinear LS problem using inverse problem theory. Two popular approaches were considered. The first approach considers the model parameters m^* as the only independent variable and considers the wavefield u^* as a function of m^* by exactly satisfying the wave equation, that is, $u^* = A(m^*)^{-1}b^*$ [5, 24, 29]. This reduced-space approach has the advantage of a relatively small problem size because the model parameters are independent of source and time. However, the associated formulation can be ill-conditioned. The interested reader is referred to [9] for AL-based formulation of the approach. The second approach considers both the model parameters and the wavefield as independent optimization variables. This formulation does not require precise satisfaction of the wave equation at each iteration of the optimization process. Instead, it only needs to be satisfied in the final solution [1, 7, 25, 36]. This formulation can act as a form of regularization, reducing the ill-conditioning of the optimization problem. However, because the wave equation is only approximately solved in each iteration, it imposes additional computational and memory burdens, which are significant disadvantages.

Following the second approach, we can estimate the model and wavefield variables simultaneously from the data by solving

$$\underset{m,u}{\text{minimize}} \ \|Pu - d^*\|_2^2 \quad \text{subject to} \quad A(m)u = b^*. \tag{3}$$

Indeed, in practical applications of FWI, regularization terms are often included in the objective function to ensure stable model updates and to impose constraints on the updated model parameters. These regularization terms help filter out undesired features and restrict the model to desired spaces [9]. However, for the purpose of interpretation and to maintain the focus on the role of the multipliers, we have chosen not to include such regularization terms in our formulas. This allows us to develop and present iterative methods in their most basic form, making it easier to analyze the contributions of the various components involved in iterations. However, it is important to note that the iterative methods presented here can easily be extended to include regularization terms. Regularization terms would introduce additional terms into the model update equations. Despite the addition of regularization, the core structure and principles of iterative methods remain unchanged, and the physical meaning of the multipliers remains important for understanding the mechanisms.

The Lagrange multiplier method formulates eq. (3) as

$$\min_{m,u} \max_v \mathcal{L}(m, u, v) = \frac{1}{2}\|Pu - d^*\|_2^2 + \langle v, A(m)u - b^* \rangle_{t,x} \tag{4}$$

where $v \in \mathbb{R}^N$ denotes the Lagrange multiplier vector and $\langle \cdot, \cdot \rangle_{t,x}$ is the canonical inner product with the integration over space and time. The KKT conditions [21], which are first-order necessary conditions for optimality of the solution, require that the partial derivatives of the Lagrangian with respect to v, u, and m be zero at the optimum point, which is a saddle point. Three sets of nonlinear equations are produced and must be solved. Newton-type methods are commonly employed to find the optimum solution, where the associated KKT system is solved simultaneously

for all variables [2, 13, 14]. In other words, a Krylov iterative method is used to solve the whole KKT system during each Newton iteration. The challenge with this method is that the Hessian matrix is very large and might be ill-conditioned and indefinite, which in turn requires appropriate preconditioning. Alternatively, this paper introduces an iterative process that breaks down the optimization problem into a saddle point (min-max) subproblem involving v and u. Subsequently, a simple preconditioned gradient descent step is performed to update the model parameters. The resulting iteration is given by:

$$(u^+, v^+) = \arg \min_{u,v} \max \mathcal{L}(m, u, v), \tag{5a}$$

$$m^+ = m - \alpha \frac{\langle v^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}. \tag{5b}$$

Here, $\alpha > 0$ represents the step length. The inner product $\langle \cdot, \cdot \rangle_t$ is performed only over time. The expression $\langle v^+, \partial_{tt} u^+ \rangle_t$ denotes multiplication and summation over time of the multiplier and the second time-derivative wave-field, which serves as the gradient of the Lagrangian function with respect to the model parameters. The denominator in the rightmost term in eq. (5b) acts as a preconditioning factor, and the division is performed element-wise.

Equation (5b) indeed offers valuable insights into the structure of the multiplier. To ensure the correct dimensionality for the model update, the multiplier should indeed involve a multiplication of the wavefield and a model perturbation. More specifically, a model update $\delta m = m^+ - m$ obtained from the evaluation of the division term in eq. (5b) shows that the multiplier should be of the form $v^+ = -\delta m \circ \partial_{tt} u^+$, where $\circ$ denotes the element-wise product in space. This form of the multiplier represents a scattering source or Born secondary source [31]. In Section 4 of this paper, we investigate this physical meaning of the multipliers.

Once the updated wavefield and multiplier are available, calculating the updated model in eq. (5b) is straightforward. Hence, all the efforts are toward developing robust methods for estimating wavefield and multiplier during the iterative process. In the following subsections, we will explore different FWI algorithms for the estimation of u^+ and v^+.

3.1 The Standard Lagrangian

We may use the original Lagrangian to calculate the updated wavefield and multiplier. The necessary conditions for optimality of them are

$$\frac{\partial \mathcal{L}(m, u, v)}{\partial v} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}(m, u, v)}{\partial u} = 0, \tag{6}$$

which result in the following two-by-two block linear system, so called saddle point system,

$$\begin{pmatrix} A(m) & 0 \\ P^T P & A(m)^T \end{pmatrix} \begin{pmatrix} u^+ \\ v^+ \end{pmatrix} = \begin{pmatrix} b^* \\ P^T d^* \end{pmatrix}. \tag{7}$$

The first equation is independent of the multiplier, which allows the wavefield to be calculated by solving the wave equation. The multiplier is then obtained by solving the adjoint equation. This results in the following iteration:

$$u^+ = A(m)^{-1} b^*, \tag{8a}$$

$$v^+ = A(m)^{-T} P^T (d^* - P u^+), \tag{8b}$$

$$m^+ = m - \alpha \frac{\langle v^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}. \tag{8c}$$

These equations define a wavefield and a multiplier that can be used as implicit functions of m in the Lagrangian function, eq. (4), reducing it to a function involving only m as in eq. (41). It can be shown that $\langle v^+, \partial_{tt} u^+ \rangle_t$ is indeed the gradient of the reduced function in eq. (41) and thus eq. (8c) is a preconditioned gradient descent step for its minimization [6, 23, 29, 37]. This implies that, in the standard FWI, the first two sets of KKT conditions are automatically satisfied. The main objective of the subsequent iterations is to satisfy the final condition, which is highly challenging owing to the poor conditioning of the problem.

Regarding the estimated multipliers, it is crucial to acknowledge that eq. (8b) provides only an approximate multiplier estimate with incorrect amplitudes, which can be out of range unless data residuals are small. This discrepancy arises from the fact that the modeling operator $PA(m)^{-1}$ is not unitary, and therefore, the estimated multiplier in eq. (8b) may not precisely fit the scattered data. In practice, this issue can impact the convergence and accuracy of the inversion process. However, we can address this challenge by tuning the step length α appropriately or using prior information to enforce the desired bounding of the multipliers within a certain range. Additionally, we can modify the Lagrangian formulation to include damping terms that penalize large Lagrange multiplier values, which helps stabilize the optimization process.

In the following subsections, we introduce the penalty and AL methods as means of regularizing the Lagrangian optimization process. However, our formulation differs slightly from the standard penalty and AL approaches, aiming to offer a simpler and more interpretable formulation while being more general. Notably, we retain the original Lagrange multiplier v in our formulation, in contrast to the conventional practice of eliminating it to obtain standard forms. It is important to note that our formulation do not introduce any additional computational or memory burdens to the final algorithms. On the contrary, it allows for a more intuitive interpretation and understanding of the optimization process.

3.2 Regularization by the quadratic penalty method

The quadratic penalty method can be seen as introducing a damping term in the Lagrangian function to penalize large multiplier values [10, 40]. This leads to a modified formulation of the min-max problem in eq. (5a):

$$(u^+, v^+) = \arg \min_{u,v} \max \ \mathcal{L}(m, u, v) - \frac{1}{2\mu} \|v\|_2^2. \tag{9}$$

Here, $\mu > 0$ controls the strength of the damping applied to the multiplier. This is simply zeroth-order Tikhonov regularization of the maximization problem. The minus sign of the regularization term comes from the maximization of the Lagrangian with respect to the multiplier. As μ tends towards infinity, the objective function approaches the original Lagrangian function, maintaining its properties. Conversely, as μ approaches zero, the multipliers dampen towards zero. By adjusting μ, we can fine-tune the influence of the damping term and enforce the desired bounding of the multipliers within a certain range.

This approach is closely connected to the standard quadratic penalty method, but it offers a simpler and more interpretable formulation. Note that in the maximization step, the Lagrange multiplier v can be explicitly computed as $v = \mu(A(m)u - b^*)$. Substituting this expression into the objective function (9) eliminates the Lagrange multiplier, simplifying the formulation to a standard quadratic penalty function in eq. (43). Also, we can eliminate u from this penalty function yielding a function solely dependent on m [7, 25, 28, 34, 35].

The damped min-max problem in eq. (9) is quadratic in both u and v variables and its optimum solution is given by the following saddle point system:

$$\begin{pmatrix} A(m) & -\frac{1}{\mu}I \\ P^T P & A(m)^T \end{pmatrix} \begin{pmatrix} u^+ \\ v^+ \end{pmatrix} = \begin{pmatrix} b^* \\ P^T d^* \end{pmatrix}. \tag{10}$$

Unlike eq. (7), this linear system is more difficult to solve. It can be solved using various standard numerical linear algebra concepts, leading to different yet equivalent iterative procedures. We can use Schur complement to write its solution explicitly as

$$u^+ = \left(P^T P + \mu A^T A\right)^{-1} \left(P^T d^* + \mu A^T b^*\right), \tag{11a}$$

$$v^+ = \left(I + \frac{1}{\mu} A^{-T} P^T P A^{-1}\right)^{-1} A^{-T} P^T (d^* - P A^{-1} b^*) \tag{11b}$$

$$= A^{-T} P^T \left(I + \frac{1}{\mu} P A^{-1} A^{-T} P^T\right)^{-1} (d^* - P A^{-1} b^*). \tag{11c}$$

where $A \equiv A(m)$. The equality of equations (11b) and (11c) can be found in Guttman [12]. However, while both equations are equivalent, eq. (11b) can be more challenging to compute due to the need for the inversion of $(I + \frac{1}{\mu} A^{-T} P^T P A)$, which is of size $N \times N$. On the other hand, eq. (11c) involves the inversion of a significantly smaller matrix $(I + \frac{1}{\mu} P A^{-1} A^{-T} P^T)$ of size $M \times M$.

Due to the complexity of matrix inversion in both cases, it is possible to compute either u^+ or v^+ explicitly and then use the first equation in eq. (10) to compute the other variable. This approach leads to two different strategies for solving the equations: the wavefield-oriented approach and the multiplier-oriented approach. In the wavefield-oriented approach, we first compute u^+ explicitly using eq. (11a), and then we utilize the first equation in eq. (10) to solve for v^+. Conversely, in the multiplier-oriented approach, we initially compute v^+ explicitly using either equation (11b) or (11c). Subsequently, we employ the first equation in eq. (10) to determine u^+.

3.2.1 Wavefield-oriented penalty method

The wavefield-oriented penalty method results in the following iteration:

$$u^+ = \left(P^T P + \mu A(m)^T A(m)\right)^{-1}(P^T d^* + \mu A(m)^T b^*), \tag{12a}$$

$$v^+ = \mu(A(m)u^+ - b^*), \tag{12b}$$

$$m^+ = m - \alpha \frac{\langle v^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}. \tag{12c}$$

3.2.2 Multiplier-oriented penalty method

The multiplier-oriented penalty method results in the following iteration:

$$v^+ = A^{-T} P^T \left(I + \frac{1}{\mu} P A^{-1} A^{-T} P^T\right)^{-1} (d^* - P A^{-1} b^*), \tag{13a}$$

$$u^+ = A^{-1}\left(b^* + \frac{1}{\mu} v^+\right), \tag{13b}$$

$$m^+ = m - \alpha \frac{\langle v^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}. \tag{13c}$$

where $A \equiv A(m)$.

3.2.3 Wavefield-oriented vs. multiplier-oriented approaches

Both wavefield-oriented and multiplier-oriented approaches are two alternative (equivalent) strategies for simultaneously computing the wavefield and multiplier, solving a saddle point system arising from Lagrangian-based FWI. The wavefield-oriented approach emphasizes wavefield reconstruction as its primary focus, while in the multiplier-oriented approach the vector of Lagrange multipliers is computed first. Each approach offers distinct perspectives and trade-offs in terms of computational efficiency and complexity. The choice between the two depends on the specific requirements and constraints of the problem at hand.

In the wavefield-oriented approach, the main challenge lies in inverting the associated Hessian matrix $(P^T P + \mu A^T A)$ for updating the wavefield. In the context of time domain applications, this matrix can be difficult to invert due to the lack of a suitable time-stepping algorithm [7, 17, 25, 38]. For small to medium scale problems in frequency domain, direct factorization methods like LU triangular factorization can be used to invert the matrix efficiently when applied to multiple sources. However, for large-scale applications, iterative methods like CG become more feasible, although appropriate preconditioning is necessary to handle the severe ill-conditioning of the Hessian matrix.

On the other hand, the multiplier-oriented approach offers the advantage of computing the wavefield using standard time-stepping methods, as long as the multiplier is available. In this approach, the computational bottleneck lies in the computation of the LS multipliers, which involves inverting the dense Hessian matrix $(I + \frac{1}{\mu} P A^{-1} A^{-T} P^T)$. However, the size of this matrix is the same as the data size [see Appendix A in 7]. While the explicit construction of this matrix requires solving N_r adjoint wave equations, its direct inversion is possible only for frequency domain applications. In time domain, approximate solutions can be obtained using iterative solvers such as preconditioned CG. In this case, it is not necessary to explicitly build the Hessian matrix but only to compute the product of it by a vector, which incurs the cost of two additional wave-equation solves per CG iteration.

3.2.4 Connection with WRI

The iteration (12) simplifies into the WRI method proposed by van Leeuwen & Herrmann [36] when the step length α is set as the inverse of the penalty parameter, i.e., $\alpha = 1/\mu$. By substituting the expression for the multiplier v^+ from eq. (12b) into the model update formula in eq. (12c), we derive a two-step WRI iteration as shown in eq. (44). However, it is important to note that WRI inherently does not satisfy the wave equation constraint. Therefore, the penalty parameter μ needs to be progressively increased during iterations to approach convergence, typically by employing a parameter continuation strategy. Yet, this strategy leads to an exacerbated ill-conditioning issue, which is a fundamental limitation of the WRI method. This arises from the fact that a larger μ results in a smaller step length, causing the algorithm to converge slowly.

This issue can be more effectively demonstrated using the formulation (13), which is equivalent to (12). By choosing $\alpha = 1/\mu$ and introducing a change of variable $\lambda = \frac{1}{\mu} v$ in eq. (13), we arrive at the following multiplier-oriented

variant of WRI:

$$\lambda^+ = A^{-T}P^T \left(\mu I + P A^{-1} A^{-T} P^T\right)^{-1}(d^* - P A^{-1} b^*), \tag{14a}$$

$$u^+ = A^{-1}(b^* + \lambda^+), \tag{14b}$$

$$m^+ = m - \frac{\langle \lambda^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}, \tag{14c}$$

This formulation demonstrates that the parameter μ effectively controls the amount of damping applied to the (scaled) multiplier λ^+. As the value of μ increases, the damping effect intensifies, potentially damping the multiplier to zero. This excessive damping causes ill-conditioning, which reflects in a slow convergence rate.

One way to address this challenge is by relaxing the constraint $\alpha = 1/\mu$, allowing for greater flexibility in controlling the algorithm's convergence. Moreover, a robust solution to this issue is offered by the AL formulation. The AL method operates effectively with moderate values of the penalty parameter, avoiding the need for excessively large values. This preserves algorithm well-posedness and enhances convergence behavior. Maintaining algorithm well-posedness comes at the cost of retaining and updating estimates of the Lagrange multipliers in each iteration. While this introduces increased memory requirements, practical strategies can be employed to efficiently manage this memory demand [7].

3.3 Regularization by the AL method

An excellent method for regularization of the Lagrangian problem is using the proximal regularization, that allows prior information of the multiplier to be introduced into the objective function [10, 26, 40]:

$$(u^+, v^+) = \arg \min_{u,v} \max \ \mathcal{L}(m, u, v) - \frac{1}{2\mu}\|v - w\|_2^2. \tag{15}$$

where w is a prior multiplier estimate. The physical interpretation of this strategy is simple: when maximizing the Lagrangian function, one wishes a multiplier v that is not too far from the a priori multiplier w. The constraints that the updated multiplier cannot be too far from some a priori estimate is necessary for avoiding ill-posedness and building a stable iteration [30]. In the proximal regularization, the prior estimate w varies with iteration [26]. It may be selected as the best estimation of the original multiplier at the previous iteration,

$$w^+ = w + \mu(A(m^+)u^+ - b^*). \tag{16}$$

The proximal Lagrangian approach is closely related to the AL method, but it provides a more intuitive and easier-to-interpret formulation. Note that the objective function in eq. (15) is quadratic in v and has a simple solution $v = w + \mu(A(m)u - b^*)$. Substituting this expression into the objective function eliminates the primal Lagrange multiplier from the optimization variables, giving the standard form of the AL function in eq. (45).

The optimum solution of the objective function in eq. (15) is given by the following saddle point system:

$$\begin{pmatrix} A(m) & -\frac{1}{\mu}I \\ P^T P & A(m)^T \end{pmatrix} \begin{pmatrix} u^+ \\ v^+ \end{pmatrix} = \begin{pmatrix} b^* - \frac{1}{\mu}w \\ P^T d^* \end{pmatrix}. \tag{17}$$

This system is essentially the same as that in eq. (10), with the only difference being the substitution of the physical source b^* at the right-hand side in the eq. (10) with the extended source $b^* - \frac{1}{\mu}w$ in eq. (17). Therefore, we can employ the two approaches presented above to solve it.

3.3.1 Wavefield-oriented AL method

Applying this approach gives us the following iteration:

$$u^+ = \left(P^T P + \mu A(m)^T A(m)\right)^{-1} \left(P^T d^* + \mu A(m)^T b^* - A(m)^T w\right), \tag{18a}$$

$$v^+ = w + \mu(A(m)u^+ - b^*), \tag{18b}$$

$$m^+ = m - \alpha \frac{\langle v^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}, \tag{18c}$$

$$w^+ = w + \mu(A(m^+)u^+ - b^*). \tag{18d}$$

If we employ the step length $\alpha = 1/\mu$ in eq. (18c) to update the model, then this iteration will be equivalent to IR-WRI [1]. IR-WRI faces a similar challenge as WRI regarding the wavefield reconstruction step, eq. (18a).

3.3.2 Multiplier-oriented AL method

The multiplier-oriented approach solves the AL problem by the following iteration:

$$v^+ = A(m)^{-T} P^T \left(I + \frac{1}{\mu} P A(m)^{-1} A(m)^{-T} P^T \right)^{-1} (d^* - PA(m)^{-1} b^* + \frac{1}{\mu} PA(m)^{-1} w), \tag{19a}$$

$$u^+ = A(m)^{-1}(b^* + \frac{1}{\mu} v^+ - \frac{1}{\mu} w), \tag{19b}$$

$$m^+ = m - \alpha \frac{\langle v^+, \partial_{tt} u^+ \rangle_t}{\langle \partial_{tt} u^+, \partial_{tt} u^+ \rangle_t}, \tag{19c}$$

$$w^+ = w + \mu(A(m^+)u^+ - b^*). \tag{19d}$$

3.3.3 Penalty vs. AL methods

In comparison to the penalty-based methods the AL-based iterations involve an additional secondary Lagrange multiplier w. This multiplier plays a crucial role in enhancing the convergence of the algorithm. Specifically, in AL-based iterations, the wave equation depicted in eq. (19b) incorporates three distinct types of sources. The known physical source, b^*, serves as the primary source, generating the background wavefield. The Lagrange multiplier v acts as a secondary volume source, contributing to the low-order scattering part of the wavefield. The Lagrange multiplier w acts as a tertiary volume source, correcting the defects in v and contributing to the high-order scattering part of the wavefield. This tertiary volume source, with its significant role in conditioning the iteration, is absent in penalty-based iterations.

The penalty-based iterations have been extensively analyzed in the framework of WRI by [22], but the numerical results were produced using the AL iteration, including this tertiary volume source term. Although there are similarities and relationships between the two approaches, they are not the same and can yield completely different results. To better comprehend the similarities and differences between these methods and the mechanisms employed by each of them in solving FWI, we need to establish the physical meaning and precisely analyze the actions of each term in the optimization process. In the subsequent section, we will provide a deeper understanding of the physical meaning of both v and w and their roles in improving the inversion results within a general framework that reveals the interconnections between different presented methods.

4   The basic iterations, mechanisms, and physical interpretations

In this section, the Lagrangian based algorithms discussed in the previous section are further developed and examined using basic iterative linearization methods and scattering concepts. Using this formulation, we gain a deeper physical understanding of the role of Lagrange multipliers in the reconstruction process. We can derive the same algorithms obtained through Lagrangian formulation but from a scattering point of view and simple alternating iterative methods. This offers a unique perspective that connects the optimization theory with the physics of wave scattering. We present the iterations in a progressive manner, starting from their simplest form and gradually advancing towards more sophisticated and refined versions. Throughout the development, we discuss the advantages and disadvantages of each iteration, highlighting their evolution and improvements. Importantly, we establish clear connections between the generated iterations and their corresponding counterparts in Lagrangian optimization, whenever necessary. By taking this comprehensive approach, we enable a thorough comparison and evaluation of the various Lagrangian based methods discussed earlier. This analysis allows us to assess their performance, understand their underlying mechanisms, and identify their strengths and limitations from the scattering theory perspective.

Let us decompose the true model m^* into the known background model m and the unknown perturbation $m^* - m$, then we can split the nonlinear operator $A(m^*)u^*$ into linear and bilinear terms, $A(m)u^*$ and $\varphi(u^*, m^* - m)$, respectively, as

$$A(m^*)u^* = A(m)u^* - \varphi(u^*, m^* - m), \tag{20}$$

[27]. Accordingly, eq. (2a) can be written in the form of the Lippmann–Schwinger integral equation [18]

$$A(m)u^* = b^* + \varphi(u^*, m^* - m). \tag{21}$$

The bilinear function φ, defined as

$$\varphi(u^*, m^* - m) = \left(-\frac{\partial A(m)}{\partial m} u^* \right)(m^* - m), \tag{22}$$
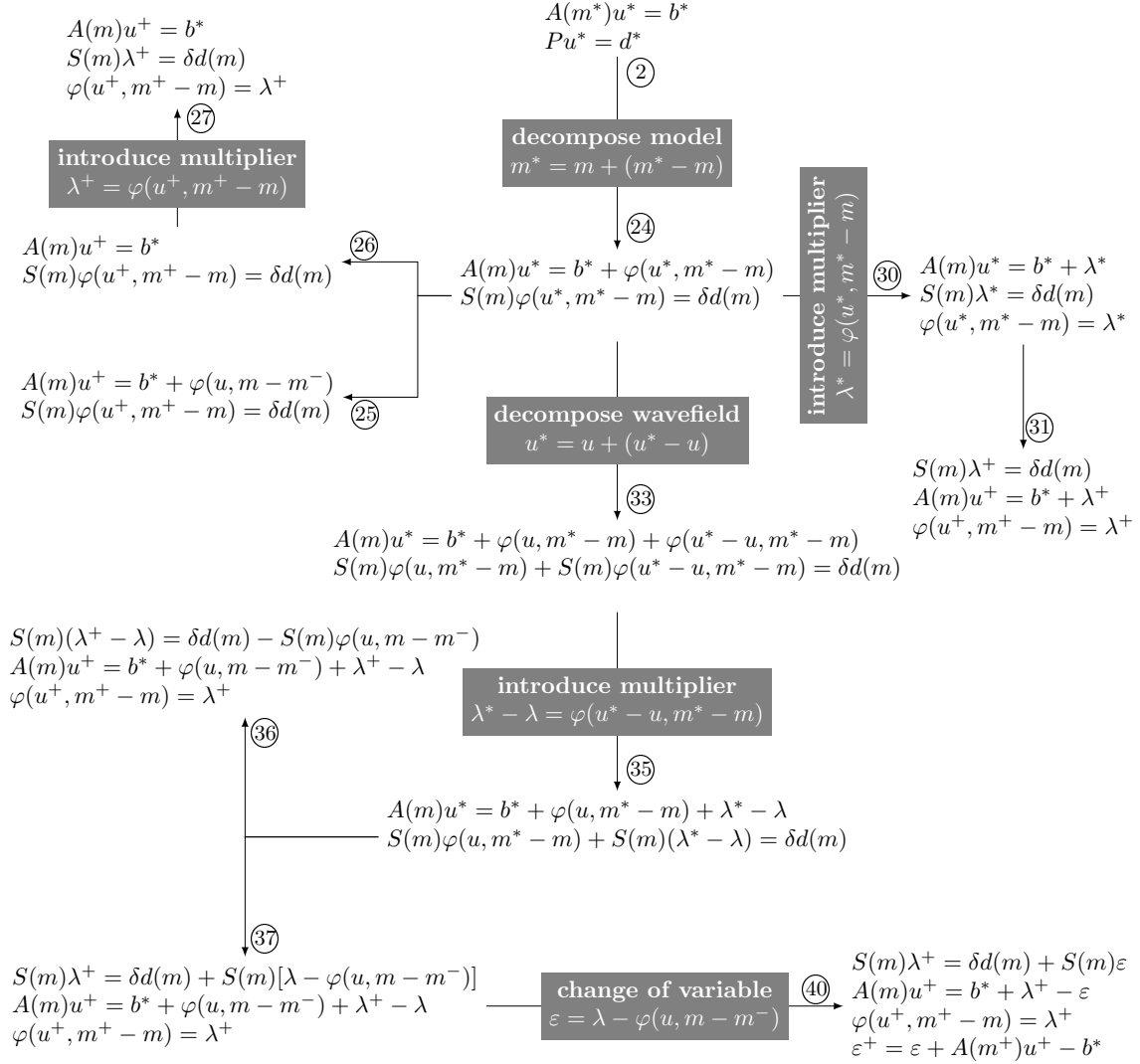
Figure 1: Interconnections between iterative methods for solving FWI equations. The circled numbers represent the corresponding equation numbers in the text. Each iteration consists of solving a series of equations sequentially, updating the variables denoted with $\bullet^+$ using the background values $\bullet$ and the previous values $\bullet^-$.

is the scattering source, also known as "secondary Born source" [31] or "contrast source" [33]. It is created by product of the perturbation, $m^* - m$, and the wavefield, and as a result, it may extend the entire space-time domain. When both arguments are considered simultaneously, φ is nonlinear, but when only one argument is considered, it is linear.

The following data equation connects the scattered data or data residual $\delta d(m) = d^* - PA(m)^{-1}b^*$ to the scattering source linearly by combining eq. (2b) and eq. (21):

$$S(m)\varphi(u^*, m^* - m) = \delta d(m). \tag{23}$$

where $S(m) = PA(m)^{-1}$. This equation represents a forward modeling operator that calculates the scattered data resulting from a given scattering source φ. However, it is still nonlinear due to the dependence of the incident wavefield u^* on the model parameters. We can linearize the equation around the background model m by treating the wavefield as an independent variable. The accuracy of this linearization is influenced by the accuracy of the wavefield, and an exact linearization can be achieved if the exact wavefield is known.

We use the splitting above to yield a new pair of coupled equations, as given in eq. (24), which are alternatives to the equations in eq. (2).

$$A(m)u^* = b^* + \varphi(u^*, m^* - m), \tag{24a}$$

$$S(m)\varphi(u^*, m^* - m) = \delta d(m). \tag{24b}$$

It should be noted that the two sets of equations in (2) and (24) are equivalent. This means that the optimal solution (m^*, u^*) obtained from either set of equations also optimally solves the other set. The coupled equations in (24) serve as the foundation to develop a variety of alternating iterations. As we will show in what follows, a number of interesting iterations, including the above Lagrangian-based algorithms, can be derived from these equations (see Figure 1). The number and type of chosen variables, as well as the linear equations between them, differ. The generated iterations may be equivalent in that they arrive at the same solution. However, the solution process may be significantly affected by the chosen formulation.

4.1 The Gauss-Seidel iteration

The coupled equations presented in eq. (24) pose a challenge for simultaneous solution due to their inherent nonlinearity. To overcome this challenge, an iterative decoupling approach can be employed. This method breaks down the problem into subproblems, with each subproblem focusing on updating only one variable by solving a single equation using information from the previous iteration.

Let us start with an initial point (m_0, u_0), set $m_{-1} = m_0$, and solve eq. (24a) for the new wavefield u_1 using $\varphi(u_0, m_0 - m_{-1})$ as an estimate of the scattering source. We determine a new model m_1 such that $\varphi(u_1, m_1 - m_0)$ solves eq. (24b) after obtaining the new wavefield u_1. Subsequently, with (m_1, u_1), the process is repeated to obtain (m_2, u_2). This Gauss-Seidel procedure is repeated until convergence is achieved. The general form of the iteration is as follows.

$$A(m)u^+ = b^* + \varphi(u, m - m^-), \tag{25a}$$

$$S(m)\varphi(u^+, m^+ - m) = \delta d(m). \tag{25b}$$

The concept behind this iteration is straightforward. By utilizing an approximate wavefield and model perturbation, we can construct an approximate scattering source in the Lippmann–Schwinger equation. A natural choice is to use the most recent estimates, $\varphi(u, m - m^-)$, which leads to an approximation of the wavefield u^+ using eq. (25a) and then model m^+ using eq. (25b).

In the Gauss-Seidel iteration, the nonlinearity of the problem is effectively addressed through two key factors. Firstly, the model is updated at each iteration, allowing for the gradual inclusion of scattering terms into the model parameters. Secondly, it accounts for remaining multiscattering effects encoded in $\varphi(u, m - m^-)$, which enables the iteration to capture higher-order scattering contributions, akin to a Born series solution. By employing the previous estimates of the wavefield and perturbation, this iterative approach represents a higher-order approximation, resulting in an enhanced estimation of the wavefield and subsequently, the model parameters. This capability allows for better convergence and improved inversion results.

However, there are two significant challenges to implementing the Gauss-Seidel iteration effectively. The first challenge pertains to computational costs, as solving eq. (25b) for the updated model is very challenging due to the large size of the associated kernel matrix which is source dependent. The second challenge lies in memory requirements. Storing and accessing the previous wavefield estimates can lead to substantial memory usage, which may hinder the applicability of the method to large-scale problems.

4.2 The Gauss-Newton iteration

Even though the Gauss-Newton based FWI is particularly recognized as a gradient based iteration for minimization of the reduced objective function in eq. (41) [19, 24], the resulting iteration can simply be obtained from the Gauss-Seidel iteration, eq. (25), by omitting the scattering source term φ from the right-hand side of eq. (25a):

$$A(m)u^+ = b^*, \tag{26a}$$

$$S(m)\varphi(u^+, m^+ - m) = \delta d(m). \tag{26b}$$

The wavefield is first computed by solving the wave equation, eq. (26a), with the physical source, starting from the background model m. We then solve the data equation (26b) to update the model using the computed wavefield. Gauss-Newton method addresses the nonlinearity of the problem only through iteration by updating the background model. The trick used in this iteration substantially reduces the memory requirements associated with storage of the

wavefields; however, it suffers from approximations in the eq. (24a) of replacing the scattering source with zero. In the case of a strong multiple scattering effect, such an approximation is inaccurate which may cause instability. However, when the background model is sufficiently accurate, this term is small and this approximation can be acceptable.

4.3 The Split Gauss-Newton iteration

The Gauss-Newton method encounters a significant computational challenge when solving eq. (26b). Directly solving this linear system can be computationally demanding and may not be practical for large-scale problems. To overcome this issue, several strategies can be employed. One option is to use an iterative method, such as the CG method, to approximate the solution of eq. (26b) [19]. Another approach to address the computational challenge is to take advantage of the split structure of the kernel matrix and decompose the solution procedure into two sequential steps, using the same strategy as solving a linear system of equations by LU decomposition [8]. In this alternative approach, we introduce an auxiliary variable (multiplier) $\lambda^+ = \varphi(u^+, m^+ - m)$, and then we split eq. (26b) into the multiplier equation, eq. (27b), and constitutive equation, eq. (27c), which are solved in sequence, leading to the split Gauss-Newton iteration:

$$A(m)u^+ = b^*, \tag{27a}$$

$$S(m)\lambda^+ = \delta d(m), \tag{27b}$$

$$\varphi(u^+, m^+ - m) = \lambda^+. \tag{27c}$$

Using the LS approach to solve the constitutive equation gives

$$\underset{m^+}{\text{minimize}} \, \|\varphi(u^+, m^+ - m) - \lambda^+\|_2^2 \quad \Longrightarrow \quad m^+ = m - \frac{\langle \lambda^+, \partial_{tt}u^+ \rangle_t}{\langle \partial_{tt}u^+, \partial_{tt}u^+ \rangle_t}. \tag{28}$$

We can see that the model update in this case is very similar to that in the Lagrangian based methods, eq. (5b). The step length here is unity and v^+ is replaced with λ^+. Interestingly, when we examine the iteration eq. (27), we can notice that it can be reduced to the Lagrangian-based iteration in eq. (8) if we estimate the multipliers by using an adjoint operator to approximately solve the multiplier equation, eq. (27b). The step length μ in eq. (8) partially accounts for these approximations. In this context, the Lagrange multipliers λ^+ and v^+ are related by the equation $\lambda^+ = (1/\mu)v^+$.

4.4 The Split Gauss-Seidel iteration

By introducing the auxiliary variable

$$\lambda^* = \varphi(u^*, m^* - m), \tag{29}$$

we can split the equations in eq. (24) as

$$A(m)u^* = b^* + \lambda^*, \tag{30a}$$

$$S(m)\lambda^* = \delta d(m), \tag{30b}$$

$$\varphi(u^*, m^* - m) = \lambda^*. \tag{30c}$$

If we reverse the order of the first and second equations and perform an alternating solve, we get eq. (31), a split form of the Gauss-Seidel iteration in eq. (25).

$$S(m)\lambda^+ = \delta d(m), \tag{31a}$$

$$A(m)u^+ = b^* + \lambda^+, \tag{31b}$$

$$\varphi(u^+, m^+ - m) = \lambda^+. \tag{31c}$$

In this iteration, the optimum λ^+ is the true scattering source, which is an optimum solution of eq. (31a). Thus, if matrix S is invertible, we can compute λ^* precisely from this equation in Step 1. Then, in step 2, the true wavefields and at step 3 converge to the true model in a single iteration. The main difficulty is that S is not invertible, which demands a more complicated iterative process to solve the problem. First, the multiplier equation (31a) is solved for the updated scattering source λ^+ and then the wavefield is obtained by solving eq. (31b) using λ^+. Finally, model m^+ is recovered by factoring the scattering source estimate into the product of the wavefield and model perturbation. Since A is invertible we can use the expression of the multiplier from eq. (31b) in eq. (31c) to get $\varphi(u^+, m^+ - m) = A(m)u^+ - b^*$. Therefore, in essence, the split Gauss-Seidel iteration, eq. (31), updates the model parameters using the Lippmann–Schwinger equation unlike the Gauss-Seidel iteration in eq. (25) which uses the data equation for this update.

This split iteration may have the following advantages: (1) it does not require storage of the previous wavefields. (2) It can be more efficient for inversion of multiple source data using the same spread acquisition, because we only need

to construct and invert the dense matrix S once and reuse it for various sources because it is source-independent. (3) The bilinear operator csc and the matrix $boldS$ have different conditioning; the former is rank defficient and may need strong regularization, whilst the latter is well posed thus inversion in the split form permits more flexibility in the suitable regularization performed at each level.

The split Gauss-Seidel iteration can be equivalent to the multiplier-oriented penalty method in eq. (13). This equivalence arises for the case $\alpha = 1/\mu$ and when we solve the multiplier equation in eq. (31a) using a damped LS approach with a damping parameter μ. In this context, the associated Lagrange multipliers are related by the equation $\lambda^+ = (1/\mu)v^+$. This equivalence provides evidence that the quadratic penalty methods like WRI utilizes a first-order Lagrange multiplier, which represents the scattering source that best fits the scattered data in a LS sense.

4.5　Approximations and errors

The iterations described by eqs. (25), (26), (27) and (31) exhibit several sources of error that are not adequately considered. It is important to recognize and address these errors to ensure accuracy and reliability. Some notable errors include the following.

1. Approximation error of the estimated scattering sources. The estimation of the scattering source involves certain approximations and assumptions. It is crucial to quantify and minimize these errors to obtain more reliable results.

2. Regularization errors. Regularization techniques are employed to stabilize the inversion process and handle ill-posed equations during updating of the model and multipliers. However, regularization introduces additional errors due to the choice of regularization parameters and assumptions regarding the statistics of the solution.

3. Simplifications in solving the equations: The iterative process often involves simplifications and assumptions in solving the system of equations involved in updating the model and multipliers. These simplifications may neglect certain terms or assume certain conditions that can lead to inaccuracies in the solutions.

In the following section, we describe refined iterations that appropriately account for these errors.

4.6　Accounting for high order scattering terms and errors

Let us decompose the true wavefield u^* into the incident wavefield u plus the scattered wavefield $u^* - u$. It should be noted that u is not necessarily the background wavefield generated by the physical source in the background model. Indeed, this is a low-order approximation of the true wavefield u^*. Thus, it may satisfy the wave equation in m only approximately, that is, $A(m)u \approx b^*$. If u is the background wavefield satisfying $A(m)u = b^*$, then $u^* - u$ is the total scattered wavefield; otherwise, it captures only the high-order part of the scattered field.

This decomposition allows us to separate the dominant (low-order) scattering term from the residual (high-order) terms as follows:
$$\varphi(u^*, m^* - m) = \varphi(u, m^* - m) + \varphi(u^* - u, m^* - m), \tag{32}$$
where $\varphi(u, m^* - m)$ represents the dominant part of the scattering source and $\varphi(u^* - u, m^* - m)$ represents the change in φ due to the error in the wavefield. Using this equation in eq. (24), we create another set of equations to be solved by alternating iterations:
$$A(m)u^* = b^* + \varphi(u, m^* - m) + \varphi(u^* - u, m^* - m), \tag{33a}$$
$$S(m)\varphi(u, m^* - m) + S(m)\varphi(u^* - u, m^* - m) = \delta d(m). \tag{33b}$$

In this form, the data residuals are decomposed into a low-order scattered data component (the first term in eq. (33b)) and the remaining multiscattered part (the second term in eq. (33b)).

Decomposition of the wavefield also allows us to define a perturbation for the Lagrange multiplier:
$$\lambda^* - \lambda = \varphi(u^* - u, m^* - m). \tag{34}$$

Utilizing eq. (34) in eq. (33), we obtain the following equations:
$$A(m)u^* = b^* + \varphi(u, m^* - m) + \lambda^* - \lambda, \tag{35a}$$
$$S(m)\varphi(u, m^* - m) + S(m)(\lambda^* - \lambda) = \delta d(m). \tag{35b}$$

To solve these equations using alternating iterations, the following steps are followed, requiring λ, u, and m^-:

1. Approximate the model perturbation $m^* - m$ in the scattering source with the previous value, $m - m^-$.

2. Estimate the multiplier λ^+ by solving eq. (35b).

3. Use these estimates to construct the scattering source and solve eq. (35a) to estimate the incident wavefield u^+.

4. Finally, generate the updated model by solving the constitutive equation using λ^+ and u^+.

The iterative process can be summarized by the following equations:

$$S(m)(\lambda^+ - \lambda) = \delta d(m) - S(m)\varphi(u, m - m^-), \tag{36a}$$

$$A(m)u^+ = b + \varphi(u, m - m^-) + \lambda^+ - \lambda, \tag{36b}$$

$$\varphi(u^+, m^+ - m) = \lambda^+. \tag{36c}$$

It begins by subtracting the scattered data $S(m)\varphi(u, m - m^-)$, computed using the most recent values of m^-, m, and u, from the data residual. This subtraction isolates the residual portion that the algorithm was unable to predict in the previous iteration, as shown in (36a). The differential term $\lambda^+ - \lambda$ is then used to best fit this remaining multiscattered portion of the data, effectively approximating the corresponding source. Therefore, this iteration can be viewed as estimating the residual multiscattering source by utilizing the multiplier equation and subsequently adding it to the recently estimated scattering source. It can be considered an improved version of iteration (25), as it considers the errors present in that iteration. Furthermore, unlike iteration (31), where the total scattering source is estimated as a whole in each iteration by solving $S(m)\lambda^+ = \delta d(m)$, this method primarily focuses on estimating the scattering residuals. This is possible because we already have a reliable estimation of the scattering source from the previous iteration, which allows us to refine and improve the solution iteratively.

Additionally, we can rearrange the terms in eq. (36) to obtain a new interpretation as a refined version of the iteration (31), as shown in eq. (37).

$$S(m)\lambda^+ = \delta d(m) + S(m)[\lambda - \varphi(u, m - m^-)], \tag{37a}$$

$$A(m)u^+ = b + \lambda^+ + \varphi(u, m - m^-) - \lambda, \tag{37b}$$

$$\varphi(u^+, m^+ - m) = \lambda^+. \tag{37c}$$

In this form, the term $\lambda - \varphi(u, m - m^-)$ captures the proximity in approximating the scattering source in the previous iteration. There may be residual errors in splitting the multiplier into the model perturbation and the wavefield. These errors in the scattering source estimate are then added back to the data residual before estimating a new multiplier.

Thus, this refined iteration combines elements from the Gauss-Seidel iteration and its split variant, in eqs. (25) and (31), carefully addressing the encountered errors and focusing on refining the scattering sources. The main advantage of this refined iteration is that upon convergence, where $\|\lambda^+ - \lambda\| = 0$, we obtain the relationship $S(m)\varphi(u, m - m^-) = \delta d(m)$. This remarkable result allows us to approximately solve eqs. (37a) and (37c) by utilizing regularized formula, while still achieving convergence towards the solution of the inverse problem.

However, it is important to note that the iterations described by eqs. (36) and (37) are primarily presented for interpretation purposes and may not be the most efficient for implementation in practice. One limitation is the requirement to store the previous values of both u and λ in memory. This can be problematic, especially for large-scale problems where memory usage is a concern. This issue can be partially addressed through a variable change, which leads to a different interpretation of the iteration. Let us introduce the variable ε as the difference between the Lagrange multiplier λ and scattering term $\varphi(u, m - m^-)$,

$$\varepsilon = \lambda - \varphi(u, m - m^-). \tag{38}$$

In fact, ε represents the error in the estimated scattering source term, including the high-order scattering part that was not predicted as well as other errors in solving the equations (see Section 4.5). By using equation eq. (37b) we can estimate the error term ε^+ in the updated variables:

$$\varepsilon^+ = \lambda^+ - \varphi(u^+, m^+ - m) = \varepsilon + A(m^+)u^+ - b^*. \tag{39}$$

Accordingly, we obtain an equivalent four-step iteration in terms of the error variable:

$$S(m)\lambda^+ = \delta d(m) + S(m)\varepsilon, \tag{40a}$$

$$A(m)u^+ = b^* + \lambda^+ - \varepsilon, \tag{40b}$$

$$\varphi(u^+, m^+ - m) = \lambda^+, \tag{40c}$$

$$\varepsilon^+ = \varepsilon + A(m^+)u^+ - b^*. \tag{40d}$$

In this iteration, the only variable that must be stored in memory is the scattering source error ε. This modification reduces the memory requirements by 50% compared with previous iteration. This four-step iteration can be interpreted as a refined version of the iteration in eq. (31). At each iteration, the residual term ε is back projected to the scattered data. Consequently, the scattered data are corrected before estimating the scattering source, leading to an improvement in the final solution and stabilization of the iterative process. The iteration presented in eq. (40) can be equivalent to the multiplier-oriented AL iteration, eq. (19). This equivalence arises for the case $\alpha = 1/\mu$ and when we solve the multiplier equation in eq. (40a) using a damped LS approach with a damping parameter μ. In this context, the associated Lagrange multipliers are related by the equation $\lambda = (1/\mu)v$ and $\varepsilon = (1/\mu)w$.

To further enhance memory efficiency, an alternative approach is to introduce Lagrange multipliers in the data space rather than in the source space. By working in the data space, which typically has a significantly smaller dimension compared to the source space, we can achieve improved memory efficiency without compromising the accuracy of the iterative solution [see 7, 9].

5   CONCLUSION

The quadratic penalty and augmented Lagrangian (AL) methods offer two distinct pathways for regularizing the standard Lagrangian-based full waveform inversion (FWI) process. These methods can be implemented through two equivalent approaches: the first involves an iterative algorithm that computes least-squares (LS) wavefields, projected subsequently to the source space for obtaining multipliers; the second yields another iterative algorithm that calculates LS multipliers, subsequently employed as secondary volume sources to compute the wavefields. Both algorithms update the model by correlating the resultant wavefields and multipliers at each spatial point.

Despite ultimately arriving at the same results, these approaches provide unique perspectives and trade-offs in terms of computational efficiency and complexity. The multiplier-oriented approach offers specific advantages: (1) extending conventional FWI procedures, (2) applicable in both frequency and time domains, and (3) affording a tangible physical interpretation.

We have shown that Lagrange multipliers can be viewed as auxiliary variables, presenting scattering terms inserted into FWI to split and linearize the governing nonlinear equations. Therefore, all of the efforts are focused on creating reliable techniques for multiplier estimate that are accurate during the iterative process. This interpretation highlights the essential role that multipliers play in improving each method's algorithmic performance and makes it easier to compare them. This study close the gap between theoretical understanding and practical applicability by clarifying the physical relevance of Lagrange multipliers and their implications. Also, it presents useful ways to improve the functionality and applicability of FWI algorithms.

A   FWI method

In the standard reduced space FWI, the wavefield in obtained by solving the wave equation, $u(m) = A(m)^{-1}b$ and the multiplier is then obtained by solving the adjoint equation, $v(m) = A(m)^{-T}P^T[d - Pu(m)]$. Due to this strategy, the Lagrangian min-max problem eq. (4) is reduced to a pure minimization problem involving only m; hence, the term reduced form.

$$\min_m \frac{1}{2}\|PA(m)^{-1}b - d\|_2^2. \tag{41}$$

The gradient of this reduced objective function with respect to the model parameters is

$$\left(\frac{\partial A(m)}{\partial m}u\right)^T A(m)^{-T}P^T(d - Pu), \tag{42}$$

where $u = A(m)^{-1}b$.

B   WRI method

The quadratic penalty formulation of the constrained problem in eq. (3) is

$$\min_{m,u} \frac{1}{2}\|Pu - d\|_2^2 + \frac{\mu}{2}\|A(m)u - b\|_2^2. \tag{43}$$

where μ is the penalty parameter. The WRI solves this objective function by alternating between minimization for u and m, which leads to the following iteration [36]:

$$u^+ = \arg\min_u \frac{1}{2}\|Pu - d^*\|_2^2 + \frac{\mu}{2}\|A(m)u - b^*\|_2^2, \tag{44a}$$

$$m^+ = \arg\min_m \frac{\mu}{2}\|A(m)u^+ - b^*\|_2^2. \tag{44b}$$

C   IR-WRI method

The augmented Lagrangian formulation of the problem eq. (3) is

$$\min_{m,u} \max_w \frac{1}{2}\|Pu - d\|_2^2 + \frac{\mu}{2}\|A(m)u - b\|_2^2 + \langle w, A(m)u - b\rangle. \tag{45}$$

Solving this objective function by the alternating direction method of multipliers gives the IR-WRI [1]

$$u^+ = \arg\min_u \frac{1}{2}\|Pu - d^*\|_2^2 + \frac{\mu}{2}\|A(m)u - b^*\|_2^2 + \langle w, A(m)u - b^*\rangle, \tag{46a}$$

$$m^+ = \arg\min_m \frac{\mu}{2}\|A(m)u^+ - b\|_2^2 + \langle w, A(m)u - b^*\rangle, \tag{46b}$$

$$w^+ = w + \mu(A(m^+)u^+ - b). \tag{46c}$$

References

[1] Aghamiry, H., Gholami, A., & Operto, S., 2019. Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers, *Geophysics*, **84(1)**, R139–R162.

[2] Akçelik, V., 2002. *Multiscale Newton-Krylov methods for inverse acoustic wave propagation*, Ph.D. thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.

[3] Alkhalifah, T. & Song, C., 2019. An efficient wavefield inversion: Using a modified source function in the wave equation, *Geophysics*, **84**(6), R909–R922.

[4] Barnier, G., Biondi, E., Clapp, R. G., & Biondi, B., 2023. Full waveform inversion by model extension: theory, design and optimization, *Geophysics*, **88**(5), 1–206.

[5] Gauthier, O., Virieux, J., & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics*, **51**(7), 1387–1403.

[6] Gauthier, O., Virieux, J., & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveform : numerical results, *Geophysics*, **51**, 1387–1403.

[7] Gholami, A., Aghamiry, H. S., & Operto, S., 2022. Extended full waveform inversion in the time domain by the augmented Lagrangian method, *Geophysics*, **87**(1), R63–R77.

[8] Gholami, A., Aghamiry, H. S., & Operto, S., 2022. On the connection between WRI and FWI: Analysis of the nonlinear term in the hessian matrix, in *SEG Technical Program Expanded Abstracts 2022*, pp. 1–5, Society of Exploration Geophysicists.

[9] Gholami, A., Aghamiry, H. S., & Operto, S., 2023. Multiplier waveform inversion (MWI): A reduced-space FWI by the method of multipliers, *Geophysics*, **88**(3), R339–R354.

[10] Gill, P. E. & Robinson, D. P., 2012. A primal-dual augmented lagrangian, *Computational Optimization and Applications*, **51**(1), 1–25.

[11] Golub, G. & Pereyra, V., 2003. Separable nonlinear least squares: the variable projection method and its applications, *Inverse problems*, **19**(2), R1.

[12] Guttman, L., 1946. Enlargement methods for computing the inverse matrix, *The annals of mathematical statistics*, pp. 336–343.

[13] Haber, E., Ascher, U. M., & Oldenburg, D., 2000. On optimization techniques for solving nonlinear inverse problems, *Inverse problems*, **16**(5), 1263.

[14] Hoffmann, A., Monteiller, V., & Bellis, C., 2021. A penalty-free approach to PDE constrained optimization: application to an inverse wave problem, *Inverse Problems*, **37**(5), 055002.

[15] Huang, G., Nammour, R., & Symes, W. W., 2018. Volume source-based extended waveform inversion, *Geophysics*, **83**(5), R369–387.

[16] Li, Y. E. & Demanet, L., 2016. Full-waveform inversion with extrapolated low-frequency data, *Geophysics*, **81**(6), R339–R348.

[17] Lin, Y., van Leeuwen, T., Liu, H., Sun, J., & Xing, L., 2023. A fast wavefield reconstruction inversion solution in the frequency domain, *Geophysics*, **88**(3), R257–R267.

[18] Lippmann, B. A. & Schwinger, J., 1950. Variational principles for scattering processes. I, *Physical Review*, **79**(3), 469.

[19] Métivier, L., Brossier, R., Operto, S., & J., V., 2017. Full waveform inversion and the truncated Newton method, *SIAM Review*, **59**(1), 153–195.

[20] Métivier, L., Brossier, R., Kpadonou, F., Messud, J., & Pladys, A., 2022. A review of the use of optimal transport distances for high resolution seismic imaging based on the full waveform, *arXiv preprint arXiv:2204.08514*.

[21] Nocedal, J. & Wright, S. J., 2006. *Numerical Optimization*, Springer, 2nd edn.

[22] Operto, S., Gholami, A., Aghamiry, H. S., Guo, G., Mamfoumbi, F., Beller, S., Aghazade, K., Mamfoumbi, F., Combe, L., & Ribodetti, A., 2023. Extending the search space of full-waveform inversion beyond the single-scattering born approximation: A tutorial review, *Geophysics*, **88**(6), 1–32.

[23] Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

[24] Pratt, R. G., Shin, C., & Hicks, G. J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophysical Journal International*, **133**, 341–362.

[25] Rizzuti, G., Louboutin, M., Wang, R., & Herrmann, F. J., 2021. A dual formulation of wavefield reconstruction inversion for large-scale seismic inversion, *Geophysics*, **86**(6), R879–R893.

[26] Rockafellar, R. T., 1976. Monotone operators and the proximal point algorithm, *SIAM journal on control and optimization*, **14**(5), 877–898.

[27] Schuster, G. T., 1985. A hybrid BIE+ Born series modeling scheme: Generalized born series, *The Journal of the Acoustical Society of America*, **77**(3), 865–879.

[28] Symes, W. W., 2020. Wavefield reconstruction inversion: an example, *Inverse Problems*, **36**(10), 105010.

[29] Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.

[30] Tarantola, A., 1986. A strategy for non linear inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.

[31] Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms including elasticity and attenuation, *Pure and Applied Geophysics*, **128**, 365–399.

[32] Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, Philadelphia.

[33] van den Berg, P. M. & Kleinman, R. E., 1997. A contrast source inversion method, *Inverse Problems*, **13**(6), 1607.

[34] van Leeuwen, T., 2019. A note on extended full waveform inversion, *arXiv preprint arXiv:1904.00363*.

[35] van Leeuwen, T. & Herrmann, F., 2016. A penalty method for PDE-constrained optimization in inverse problems, *Inverse Problems*, **32(1)**, 1–26.

[36] van Leeuwen, T. & Herrmann, F. J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophysical Journal International*, **195(1)**, 661–667.

[37] Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.

[38] Wang, C., Yingst, D., Farmer, P., & Leveille, J., 2016. Full-waveform inversion with the reconstructed wavefield method, in *SEG Technical Program Expanded Abstracts*, pp. 1237–1241.

[39] Warner, M. & Guasch, L., 2016. Adaptive waveform inversion: Theory, *Geophysics*, **81**(6), R429–R445.

[40] Wright, S. J. & Recht, B., 2022. *Optimization for data analysis*, Cambridge University Press.

[41] Yang, Y., Engquist, B., Sun, J., & Hamfeldt, B. F., 2018. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, *Geophysics*, **83**(1), R43–R62.