Lift and Relax for PDE-Constrained Inverse Problems in Seismic Imaging

Zhilong Fang* and Laurent Demanet†

Abstract—We present lift and relax for waveform inversion (LRWI), an approach that mitigates the local minima issue in seismic full waveform inversion (FWI) via a combination of two convexification techniques. The first technique (Lift) extends the set of unknown variables to their products, arranged as a moment matrix. This algebraic idea is a celebrated way to replace a hard polynomial optimization problem by a semidefinite programming approximation. Concretely, both the model and the wavefield are lifted from vectors to rank-2 matrices. The second technique (Relax) invites to relax the strict wave-equation constraint—a technique known as waveform reconstruction inversion (WRI), which introduces wave-equation misfits as a weighted penalty term in the objective function. The relaxed penalty formulation enables balancing the data and wave-equation misfits by tuning a penalty parameter. Together, “Lift” and “Relax” help reformulate the inverse problem as a set of constraints on a rank-2 moment matrix. Such a lifting strategy permits good data and wave equation fits throughout the inversion process while leaving the numerical rank of the rank-2 moment matrix to be minimized down to one. Moreover, LRWI does not require adjoint wavefield to compute the gradient, which mitigates computational burdens. Numerical examples indicate that starting with a poor initial model, LRWI can conduct successful inversions with a starting frequency that is higher than that required by FWI and WRI.

Index Terms—Geophysical data, surface and subsurface properties.

I. INTRODUCTION

Seismic imaging is the primary means for Earth scientists and geophysicists to explore and study Earth’s deep interior, where direct observations are infeasible. Recently, full waveform inversion (FWI) [1]–[3] has become one of the most important seismic imaging approaches because of its potential capability in creating high-resolution subsurface images through the usage of all kinds of waves in the data.

Conventional FWI seeks a subsurface velocity model that can minimize the difference between its predicted data and the observed data in a least-squares sense. A well-known problem associated with conventional FWI is that it suffers from local minima in the objective function caused by the so-called “cycle-skipping” issues. More specifically, if the initial model does not generate predicted data within half a wavelength of the observed data, iterative optimization approaches may stagnate at physically meaningless solutions with a high probability. To conduct a successful inversion, conventional FWI needs a good initial model that is kinematically accurate at the longest data wavelengths and data containing enough low frequencies and long offsets [3]–[5]. Research aimed at mitigating the “cycle-skipping” issue mainly focuses on different misfit functions [6]–[11], expanding the search space [12]–[15], and the integration with the advanced approach of migration velocity analysis [16], [17].

We propose a two-pronged lift and relax waveform inversion (LRWI) approach to mitigating the local minima problem in this article. The proposed approach consists of two strategies that expand the search space. The “Relax” strategy is based on the so-called approach wavefield reconstruction inversion (WRI) [12], [14]. WRI first introduces wavefields as additional unknown variables and then weakens the partial differential equation (PDE) constraints used in conventional FWI by treating the PDE misfit as a weighted penalty term in the objective function. Through tuning the penalty parameter, the resulting approach does not enforce the PDE constraints at each iteration and arguably yields a less nonlinear problem in the model parameter. The “Lift” strategy follows the early work in [18] that borrows ideas from recent developments in the semidefinite relaxation for polynomial equations to mitigate the nonconvexity [19], [20]. We lift both unknown wavefields and model parameters from 1-D vectors to rank-2 matrices and reformulate the WRI problem as a set of constraints on a rank-2 moment matrix. Such a lifting strategy permits a good data and wave-equation fit throughout the inversion process while leaving the numerical rank of the moment matrix to be the quantity to minimize—so that this matrix aims to be a rank one matrix at convergence eventually.

Compared with conventional FWI, the proposed LRWI approach has three major advantages. First, the computation of the gradients does not require adjoint or reverse-time wavefields. Second, the “Relax” and “Lift” strategies enable us to fit both data misfit and PDE misfit even with poor models. Third, the rank-2 formulation provides us with the potential to utilize information from the two components in the rank-2 model matrix simultaneously. The last two properties, in conjunction with the expanded search space, may result in an optimization formulation that is less prone to local minima.

Lifting unknown parameters from vectors to rank-2 matrices yields an increase in the computational cost. Therefore, we suggest not using LRWI for the whole inversion, but using it to create a better starting model for FWI. We present numerical examples on both Marmousi and Overthrust models to illustrate that LRWI can build better starting models for FWI using initial models and starting frequencies that FWI considers as too poor and too high, respectively.

*Zhilong Fang is with the Department of Mathematics, Massachusetts Institute of Technology, Boston, MA 02139 USA (e-mail: fzlong88@gmail.com).
†Laurent Demanet is with the Department of Mathematics, Massachusetts Institute of Technology, Boston, MA 02139 USA, and also with the Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Boston, MA 02139 USA.

Digital Object Identifier 10.1109/TGRS.2020.3037856

1558-0644 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
This article is organized as follows. First, we review the basic conception and formulation of FWI. Next, we derive the formulation for LRWI and all the necessary components for the efficient optimization strategy. Finally, we present numerical examples to illustrate the feasibility and advantages of LRWI and conclude the article with a detailed discussion.

II. METHODOLOGY

Given a data set \( \mathbf{d} \in \mathbb{C}^{n_s \times n_t \times n_f} \) with \( n_s \) sources, \( n_t \) receivers, and \( n_f \) frequencies, FWI reconstructs the discretized \( n_s \times n_t \times n_f \) dimensional squared slowness model \( \mathbf{m} \) from \( \mathbf{d} \) by solving the following PDE-constrained optimization problem:

\[
\min_{\mathbf{m}, \mathbf{u}} f_f(\mathbf{m}, \mathbf{u}) = \frac{1}{2} \sum_{i,j} \| \mathbf{P} \mathbf{u}_{i,j} - \mathbf{d}_{i,j} \|_2^2 \\
\text{s.t.} \quad \mathbf{A}_f(\mathbf{m}) \mathbf{u}_{i,j} = \mathbf{q}_{i,j}
\]

where the matrix \( \mathbf{A}_f(\mathbf{m}) = \Delta + \omega_j^2 \text{diag}(\mathbf{m}) \) is the discretized Helmholtz matrix with the Laplace operator \( \Delta \). The operator \( \mathbf{P} \) projects wavefields \( \mathbf{u}_{i,j} \) corresponding to the \( t \)th source \( \mathbf{q}_{i,j} \) with frequency \( \omega_j \) onto receiver locations. When receivers are located on computational grids, \( \mathbf{P} \) picks the components of \( \mathbf{u}_{i,j} \) whose indexes correspond to the receiver locations.

The optimization problem in (1) requires a solution in \( \mathbb{R}^{n_s \times n_t \times n_f} \) with \( n_s = n_t = n_f = n_g \), which is infeasible for most practical applications due to the huge storage cost. To reduce the storage cost, one widely used approach is the adjoint-state method [3], which eliminates the PDE constraint by solving it directly, yielding the following reduced problem:

\[
\min_{\mathbf{m}} f_r(\mathbf{m}) = \frac{1}{2} \sum_{i,j} \| \mathbf{P} \mathbf{A}_f(\mathbf{m})^{-1} \mathbf{q}_{i,j} - \mathbf{d}_{i,j} \|_2^2.
\]

Consequently, the dimensionality of the search space reduces from \( n_s + n_g \) to \( n_g \). However, the tradeoff lies in the fact that the inversion of the Helmholtz matrix introduces a very strong nonlinearity into the problem, yielding an objective function \( f_r(\mathbf{m}) \) with many local minima.

A. WRI With a Rank-1 Relaxation

We aim to mitigate the local minima issue of FWI by proposing a Lift and Relax formulation in the rank-1 case. Since we compute wavefield and objective function for each source and each frequency sequentially, we will omit the dependence of the variables on the source and frequency indexes \( i \) and \( j \) from now on to simplify the notation.

We follow [12] and relax the PDE constraint in (1) by considering the PDE misfit as a penalty term as follows:

\[
\min_{\mathbf{m}, \mathbf{u}} f_p(\mathbf{m}, \mathbf{u}) = \frac{1}{2} \| \mathbf{P} \mathbf{u} - \mathbf{d} \|_2^2 + \frac{\lambda}{2} \| \mathbf{A}(\mathbf{m}) \mathbf{u} - \mathbf{q} \|_2^2.
\]

The penalty parameter \( \lambda \) enables us to balance the PDE and data misfits and provides the freedom to design a search path in the enlarged space that can potentially bypass the local minima in the objective function of conventional FWI.

Following the PDE relaxation, we introduce an additional rank-1 relaxation to expand the search space into a higher dimension space, which is motivated from the following matrix expression of the unknown parameters \( \mathbf{m} \) and \( \mathbf{u} \):

\[
\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \mathbf{X}_{13} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & \mathbf{X}_{23} \\ \mathbf{X}_{31} & \mathbf{X}_{32} & \mathbf{X}_{33} \end{bmatrix} = [1, \mathbf{m}^T, \mathbf{u}^T]^T [1, \mathbf{m}^T, \mathbf{u}^T].
\]

Clearly, the matrix \( \mathbf{X} \) is a rank-1 positive semidefinite matrix. Based on (4), we can lift the original WRI problem from optimizing over vectors \( \mathbf{m} \) and \( \mathbf{u} \) to optimizing over the matrix \( \mathbf{X} \). In the course of doing so, the direct correspondence to \( \mathbf{m} \) and \( \mathbf{u} \) in (4) is not directly imposed, but the objective in (3) is rewritten with the blocks of \( \mathbf{X} \) serving as proxies for \( \mathbf{m} \), \( \mathbf{u} \), and the product \( \mathbf{m} \mathbf{u}^T \), i.e., replacing the vectors \( \mathbf{u} \) and \( \text{diag}(\mathbf{m}) \) in (3) by \( \mathbf{X}_{31} \) and \( \text{diag}(\mathbf{X}_{32}) \). This yields the following equivalent optimization problem:

\[
\min_{\mathbf{X}} f_{px}(\mathbf{X}) = \frac{1}{2} \| \mathbf{P} \mathbf{X}_{31} - \mathbf{d} \|_2^2 \\
+ \frac{\lambda}{2} \| \mathbf{X}_{31} + \omega^2 \text{diag}(\mathbf{X}_{32}) - \mathbf{q} \|_2^2 \\
\text{s.t.} \quad \mathbf{X}_{11} = 1, \quad \mathbf{X} \succeq 0, \quad \text{rank}(\mathbf{X}) = 1.
\]

The new objective function \( f_{px}(\mathbf{X}) \) is quadratic with respect to the matrix \( \mathbf{X} \), which is much simpler than the original FWI and WRI objective functions. Since \( \mathbf{X} \in \mathbb{C}^{(n_s + n_g + 1)^2} \), we are not able to optimize over \( \mathbf{X} \) directly for large-scale realistic applications. Nonetheless, as stated by Cosse et al. [18], it is possible for us to obtain a computationally feasible formulation with a reasonable storage requirement by introducing a rank-1 approximated factorization \( \mathbf{RR}^T \) for the matrix \( \mathbf{X} \):

\[
\min_{\mathbf{R}} f_{pr}(\mathbf{R}) = \frac{1}{2} \| \mathbf{P} (\mathbf{R}_{11})^T - \mathbf{d} \|_2^2 \\
+ \frac{\lambda}{2} \| \mathbf{A}(\mathbf{R}_{22})^T + \omega^2 \text{diag}(\mathbf{R}_{23}) - \mathbf{q} \|_2^2 \\
\text{s.t.} \quad \mathbf{R}_{11} \mathbf{R}_{22}^T = 1
\]

where \( \mathbf{R} = (\mathbf{R}_{11}, \mathbf{R}_{12}, \mathbf{R}_{22})^T \) with \( \mathbf{R}_{11} = [a_1, \ldots, a_r] \in \mathbb{R}^{1 \times r}, \mathbf{R}_{22} = [\tilde{\mathbf{m}}_1, \ldots, \tilde{\mathbf{m}}_r] \in \mathbb{R}^{n_g \times r}, \) and \( \mathbf{R}_{3} = [\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_r] \in \mathbb{C}^{n_s \times r} \). When \( r = 1 \), the optimization problem in (6) will reduce to the original WRI problem in (3). A larger \( r \) yields a stronger relaxation but introduces more computational costs and storage requirements.

B. WRI With a Rank-2 Relaxation

To balance the relaxation and computational costs, we present a rank-2 formulation for the optimization problem in (6). With \( r = 2 \), we have

\[
\mathbf{m} = a_1 \tilde{\mathbf{m}}_1 + a_2 \tilde{\mathbf{m}}_2, \quad \mathbf{u} = a_1 \tilde{\mathbf{u}}_1 + a_2 \tilde{\mathbf{u}}_2
\]

\[
\mathbf{m} \odot \mathbf{u} = \tilde{\mathbf{m}}_1 \odot \tilde{\mathbf{u}}_1 + \tilde{\mathbf{m}}_2 \odot \tilde{\mathbf{u}}_2, \quad 1 = a_1^2 + a_2^2
\]

where the operator \( \odot \) represents the pointwise product. This rank-2 expression yields the following optimization problem:

\[
\min_{\tilde{\mathbf{m}}, \tilde{\mathbf{u}}, \alpha} f_{pr}(\tilde{\mathbf{m}}, \tilde{\mathbf{u}}, \alpha) = \frac{1}{2} \left( \sum_{l=1}^{r} \| \mathbf{P} \alpha_l \tilde{\mathbf{u}}_l - \mathbf{d} \|_2^2 \\
+ \frac{\lambda}{2} \left( \sum_{l=1}^{r} \alpha_l \Delta \tilde{\mathbf{u}}_l + \omega^2 \sum_{l=1}^{r} \tilde{\mathbf{m}}_l \odot \tilde{\mathbf{u}}_l - \mathbf{q} \right) \right). \]

(8)

where \( \alpha_l \) are the weights used to balance the relaxation and the computational costs.
In (8), we aim to find the best $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{u}}$ to fit both the observed data and the PDE. However, this optimization problem has infinite solutions. Indeed, for any fixed pair of $(\tilde{\mathbf{m}}^*, \alpha^*)$, the optimal $\tilde{\mathbf{u}}^*$ for the objective function $f_{p_2}(\tilde{\mathbf{m}}^*, \tilde{\mathbf{u}}, \alpha^*)$ should satisfy $\mathbf{S}\tilde{\mathbf{u}}^* = \lbrack \mathbf{d}^T, \lambda^* \mathbf{q}^T \rbrack^T$ with
\[
\mathbf{S} = \begin{bmatrix} \alpha_1 \mathbf{P}, & \alpha_2 \mathbf{P}, \\
\lambda^* (\alpha_1 \Delta + \omega_2^2 \tilde{\mathbf{m}}_1), & \lambda^* (\alpha_2 \Delta + \omega_2^2 \tilde{\mathbf{m}}_2) \end{bmatrix}
\]
and $\tilde{\mathbf{u}}^* = \lbrack \tilde{\mathbf{u}}^T \rbrack$. Since the matrix $\mathbf{S}$ is an $(n_g \times n_r) \times 2n_g$ matrix with $n_g > n_r$, there are infinite solutions for $\tilde{\mathbf{u}}^*$, which yields infinite global minima $(\tilde{\mathbf{m}}^*, \tilde{\mathbf{u}}^*, \alpha^*)$’s satisfying $f_{p_2}(\tilde{\mathbf{m}}^*, \tilde{\mathbf{u}}^*, \alpha^*) = 0$.

To mitigate the nonuniqueness issue of optimizing (8), we design a regularization based on the constraint of rank($\mathbf{X}$) = 1 in (5). A necessary condition for rank($\mathbf{R}$) = 1 is that its three components $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ should satisfy the following requirements:
\[
a_1 \tilde{\mathbf{m}}_2 = a_2 \tilde{\mathbf{m}}_1, \quad a_1 \tilde{\mathbf{u}}_2 = a_2 \tilde{\mathbf{u}}_1, \quad \tilde{\mathbf{m}}_1 \odot \tilde{\mathbf{u}}_2 = \tilde{\mathbf{m}}_2 \odot \tilde{\mathbf{u}}_1. \] (9)

We can use these properties to regularize the problem. Since $\mathbf{m}$ and $\tilde{\mathbf{u}}$ are interested, we use the third property and obtain
\[
\begin{aligned}
\min_{\mathbf{m}, \tilde{\mathbf{u}}, \alpha} f_{p_2}(\mathbf{m}, \tilde{\mathbf{u}}, \alpha) &= \frac{1}{2} \left\| \sum_{i=1}^2 \mathbf{P} \tilde{\mathbf{u}}_i - \mathbf{d} \right\|^2_2 \\
+ \frac{\lambda}{2} \left\| a_1 \Delta \tilde{\mathbf{u}}_1 + \omega_2^2 \tilde{\mathbf{m}}_1 \odot \tilde{\mathbf{u}}_1 - \mathbf{q} \right\|^2_2 \\
+ \gamma \left\| \tilde{\mathbf{m}}_1 \odot \tilde{\mathbf{u}}_2 - \tilde{\mathbf{m}}_2 \odot \tilde{\mathbf{u}}_1 \right\|^2_2 \\
\end{aligned}
\]
s.t. $a_1^2 + a_2^2 = 1$ (10)
where $\gamma$ controls the strength of the rank-1 regularization.

Finally, we can eliminate the constraint $a_1^2 + a_2^2 = 1$ in (10) by the polar coordinates transform $a_1 = \sin \theta$ and $a_2 = \cos \theta$, yielding
\[
\begin{aligned}
\min_{\mathbf{m}, \tilde{\mathbf{u}}, \theta} f_{p_2}(\mathbf{m}, \tilde{\mathbf{u}}, \theta) &= \frac{1}{2} \left\| \mathbf{P}(\sin \theta \tilde{\mathbf{u}}_1 + \cos \theta \tilde{\mathbf{u}}_2) - \mathbf{d} \right\|^2_2 \\
+ \frac{\lambda}{2} \left\| \Delta (\sin \theta \tilde{\mathbf{u}}_1 + \cos \theta \tilde{\mathbf{u}}_2) + \omega_2^2 \sum_{i=1}^2 \tilde{\mathbf{m}}_i \odot \tilde{\mathbf{u}}_i - \mathbf{q} \right\|^2_2 \\
+ \gamma \left\| \tilde{\mathbf{m}}_1 \odot \tilde{\mathbf{u}}_2 - \tilde{\mathbf{m}}_2 \odot \tilde{\mathbf{u}}_1 \right\|^2_2 \\
\end{aligned}
\]
(11)
The rank-2 relaxation is a compromise between the convexity of the problem and the necessary computational and storage costs. Although the rank-2 relaxation weakens the convexity of the problem, the objective function in (11) enables us to simultaneously fit both data and PDEs even with poor models. As a result, the resulting approach can help us find a new search path in the high-dimensional space that mitigates local minima introduced by cycle skipping.

### C. Variable Projection and Optimization Scheme

The optimization problem in (11) still faces the challenge of a large storage requirement. We use the variable projection method [21] to project out the wavefields $\tilde{\mathbf{u}}$, which is the main source of the storage cost. For any pair of $\mathbf{m}$, the objective function $f_{p_2}(\mathbf{m}, \tilde{\mathbf{u}})$ is quadratic with respect to $\tilde{\mathbf{u}}$, whose minimizer has an analytical solution
\[
\tilde{\mathbf{u}}^* = \left( \mathbf{S}^\top \mathbf{S} \right)^{-1} \mathbf{S}^\top \lbrack \mathbf{d}^T, \lambda^* \mathbf{q}^T, 0 \rbrack^T
\]
(12)
with
\[
\mathbf{S} = \begin{bmatrix} \sin \theta \mathbf{P} & \cos \theta \mathbf{P}, \\
\lambda^* \mathbf{A}(\mathbf{m}_1) & \lambda^* \mathbf{A}(\mathbf{m}_2) \\
\gamma \mathbf{d}^\top \mathbf{S} & -\gamma \mathbf{d}^\top \mathbf{S} \end{bmatrix}
\]
\[
\mathbf{A}(\mathbf{m}) = \sin \theta \Delta + \omega_2^2 \mathbf{d}^\top \mathbf{m}, \quad \text{and} \quad \mathbf{A}(\mathbf{m}_2) = \cos \theta \Delta + \omega_2^2 \mathbf{d}^\top \mathbf{m}_2.
\]

Substituting the variable $\tilde{\mathbf{u}}$ in (11) by the optimal solution $\tilde{\mathbf{u}}^*$, we obtain a reduced objective function $\overline{f}_{p_2}(\mathbf{m}, \tilde{\mathbf{u}})$, whose minimizer has an analytical solution
\[
\mathbf{m}^* = \left( \mathbf{S}^\top \mathbf{S} \right)^{-1} \mathbf{S}^\top \lbrack \mathbf{d}^T, \lambda^* \mathbf{q}^T, 0 \rbrack^T
\]
(12)
Algorithm 1 Rank-2 LRWI

1. Initialization with $\tilde{m}_1^{(0)}$, $\tilde{m}_2^{(0)}$ and $\theta^{(0)}$

2. for $k = 1 \rightarrow n_{a}$

3. Compute $\tilde{u}^{(k)}$ by Equation (12)

4. Compute $\tilde{J}_{p_2}^{(k)}(\tilde{m}^{(k)}, \theta^{(k)})$ and $\nabla_{m_{p_1}}\tilde{J}_{p_2}^{(k)}(\tilde{m}^{(k)}, \theta^{(k)})$ by Equations 14

5. 1-BFGS step in $\tilde{m}^{(k)}$ to get $\tilde{m}^{(k+1)}$

6. Compute $\nabla_{\theta}\tilde{J}_{p_1}^{(k)}(\tilde{m}^{(k+1)}, \theta^{(k)})$ by Equations 14

7. Gradient descent step in $\theta^{(k)}$ to get $\theta^{(k+1)}$

8. end

9. Obtain $\theta^{*}$ and $\tilde{m}^{*} = (\tilde{m}_1^{*}, \tilde{m}_2^{*})$

10. Output $\mathbf{m}^{*} = \sin \theta^{*}\tilde{m}_1^{*} + \cos \theta^{*}\tilde{m}_2^{*}$

D. Selection of $\lambda$ and $\gamma$

The selection of $\lambda$ and $\gamma$ plays an important role in the proposed LRWI because they affect the condition number of the matrix $\tilde{S}$ and the search path. An appropriate selection can produce a search path that bypasses the local minima of FWI and also speeds up the optimization procedure. We propose a two-stage unit-free strategy to select $\lambda$ and $\gamma$.

According to [12] and [15], $\lambda$ can be determined by the largest eigenvalue $\mu_1$ of the matrix $A^{-1}P^TPA^{-1}$. $\lambda > \mu_1$ can be considered large, whereas $\lambda < \mu_1$ can be considered small. Considering the computational cost, we select $\lambda$ according to the value $\mu_1(A(m^{(0)})^{-1}P^TPA(m^{(0)})^{-1})$ with the initial model $m^{(0)} = \sin \theta^{(0)}m_0 + \cos \theta^{(0)}m_0$.

With $\lambda$ in hand, the selection of $\gamma$ determines the condition number of the matrix $\tilde{S}$. Since both $\begin{bmatrix} \sin \theta P & \cos \theta P \\ \frac{1}{2}\tilde{A}(\tilde{m}_1) & \frac{1}{2}\tilde{A}(\tilde{m}_2) \end{bmatrix}$ and $\begin{bmatrix} \gamma \frac{1}{2}\tilde{A}(\tilde{m}_1) & \frac{1}{2}\tilde{A}(\tilde{m}_2) \\ \frac{1}{2}\tilde{A}(\tilde{m}_1) & \frac{1}{2}\tilde{A}(\tilde{m}_2) \end{bmatrix}$ are underdetermined, either very large or small $\gamma$ will yield $\tilde{S}$ bad conditioned. Consider the block expression $\tilde{S}^T\tilde{S} = \begin{bmatrix} T_{1,1} & T_{1,2} \\ T_{2,1} & T_{2,2} \end{bmatrix}$ with $T_{i,j} = \alpha_i\alpha_jP^TP + \frac{1}{2}\tilde{A}(\tilde{m}_i)\tilde{A}(\tilde{m}_j) + (\mp i)^{\lambda+1}\gamma\delta(\tilde{m}_{i}-\tilde{m}_{j,0})\delta(\tilde{m}_{j,0})$. A natural scale for $\gamma$ is the fraction between the $\ell_2$-norm of the vector $\tilde{m}_i \odot \tilde{m}_j$, i.e., $\|\tilde{m}_i \odot \tilde{m}_j\|_2$.

At the early stage of the optimization, we can select small $\lambda$ and $\gamma$ to relax both constraints. As the optimization proceeds, we can increase $\lambda$ and $\gamma$ to strengthen both constraints so that the solution can converge to the optimal solution of FWI.

E. Computational Cost Analysis

The major computational cost is to compute the optimal wavefields $\tilde{u}^{*}$ by inverting the $2n_g \times 2n_g$ matrix $\tilde{S}^T\tilde{S}$ in (12). If we use a direct solver to invert $\tilde{S}^T\tilde{S}$, the computational cost will be $O(8n_g^2)$. With $\tilde{u}^{*}$ in hand, the computation of the gradients does not require to invert additional matrices, yielding an ignorable computational cost. Since we need to update $\tilde{m}$ and $\theta$ alternately in each iteration, the total computational cost for LRWI is $O(16n_g^2n_f)$. Compared with FWI, whose computational cost is $O(2n_f^3n_f)$, LRWI is eight times as expensive as FWI. Considering the increased computational cost, we suggest not using LRWI for the whole inversion, but using LRWI to create a better initial model for FWI.

For a rank-$r$ relaxation, the size of the matrix $\tilde{S}^T\tilde{S}$ is $rn_g \times rn_g$. As a result, the corresponding computational cost is $O(2r^3n_g^2n_f)$. Due to the cubic growth of the computational cost with respect to $r$, we are restricted to select a small $r$.

III. NUMERICAL EXAMPLES

A. Marmousi Model

We conduct an example on the Marmousi-2 model $m$, shown in Fig. 1(a) to study the feasibility of LRWI. We use a Ricker wavelet centered at 15 Hz to simulate 49 sources at the depth of 0.04 km with a sampling interval of 0.5 km. We place 247 receivers at the same depth with a sampling interval of 0.04 km. As is commonly practiced, we perform the frequency continuation [23] using three frequency bands of $[2.0, 2.5, 3.0]$ Hz, $[5.0, 6.0, 7.0]$ Hz, and $[7.0, 8.0, 9.0]$ Hz. We discretize the model with 0.04 km grids. We compare the performances of FWI, WRI, and LRWI. For FWI and WRI, we use the 1-BFGS method to solve the optimization problem, whereas we use Algorithm 1 to solve LRWI. We use LRWI to conduct an inversion on the lowest frequency band and then use the obtained model as the initial model for FWI. All three approaches use 45 iterations for each frequency band.

We start FWI and WRI with the 1-D monotonously increasing velocity model $m^{(0)}$ in Fig. 1(b). For LRWI, we select $\theta^{(0)} = \pi/4$ and $m^{(0)} = (\sin \theta^{(0)}m_0^{(0)}, \cos \theta^{(0)}m_0^{(0)})$. We conduct WRI with four different selections of $\lambda$, i.e., $\lambda = \beta_1 \mu_1(A^TP^TPA^{-1})$ with $\beta_1 = 1e-8, 1e-4, 1e0, and 1e4$. For LRWI, we use the same selection for $\lambda$ and select six different $\gamma$’s for each $\lambda$. We select $\gamma = \beta_2 \mu_2(\tilde{A}(\lambda))$, where $\beta_2$ is $1e-16, 1e-12, 1e-8, 1e-4, 1e0, and 1e4$. The selections of $\beta_1$ and $\beta_2$ cannot be extremely small; otherwise, the matrix $\tilde{S}(\beta_1, \beta_2)^T\tilde{S}(\beta_1, \beta_2)$ can be close to singular or badly scaled.

Before the inversion, we first study the condition number of the matrix $\tilde{S}^T\tilde{S}$ with respect to the selection of $\beta_1$ and $\beta_2$. We use the initial model $m^{(0)}$ to build the Helmholtz matrix $A(m^{(0)})$ and compute the condition number of the matrix $A(m^{(0)})^TA(m^{(0)})$ as a reference [see the blue line in Fig. 2(a)]. Then, we use the initial model $m^{(0)}$ and different selections of $\beta_1$ and $\beta_2$ to build the matrix $\tilde{S}(\beta_1, \beta_2)^T\tilde{S}(\beta_1, \beta_2)$. The ranges for $\beta_1$ and $\beta_2$ are $[1e-3, 1e3]$ and $[1e-8, 1e4]$, respectively. The condition number of the matrix $\tilde{S}(\beta_1, \beta_2)^T\tilde{S}(\beta_1, \beta_2)$ with respect to different selections of $\beta_1$ and $\beta_2$ is plotted in Fig. 2(a). We can observe that with the given selection of $\beta_1$ and $\beta_2$, the condition number of $\tilde{S}(\beta_1, \beta_2)^T\tilde{S}(\beta_1, \beta_2)$ is not significantly larger than that of $A(m^{(0)})^TA(m^{(0)})$. Meanwhile,
Fig. 2. (a) Condition number of the matrix $\mathbf{S}^T\mathbf{S}$ versus the values of $\beta_1$ and $\beta_2$. (b) Relative model error comparison for WRI and LRWI with different selections of $\beta_1$ and $\beta_2$.

Fig. 3. (a)–(c) Results of FWI, WRI, and LRWI after first frequency band. (d)–(f) Final results of FWI, WRI, and LRWI.

Fig. 4. (a) Comparison of the absolute data difference $|d_{\text{obs}} - d_{\text{pred}}|$ for the source located at $x = 9$ km and frequency of 3 Hz. The three lines correspond to FWI (black), WRI (blue), and LRWI (red), respectively. (b) Relative model error comparison for FWI (s), WRI (o), and LRWI (x) using data with different starting frequencies.

Fig. 5. (a) True model. (b) Initial model.

Robustness to the Starting Frequency: To investigate the robustness of the three methods with respect to the starting frequency, we conduct an experiment with the starting frequency varying from 0.5 to 3.0 Hz. We use the initial model in Fig. 1(b). Fig. 4(b) shows the relative model errors versus the starting frequency for all the three methods. The highest starting frequencies for FWI, WRI, and LRWI to obtain a result with an acceptable relative model error ($\leq 14\%$) are 1, 1, and 2 Hz, respectively. This comparison implies that under current settings, the lowest necessary starting frequency of LRWI is twice as high as that of FWI and WRI.

B. Overthrust Model

We conduct a second experiment with the Overthrust model to investigate the robustness of LRWI with respect to different velocity structures. Fig. 5(a) shows the 5 km $\times$ 20 km Overthrust model. We place 99 sources and 100 receivers at the depth of 0.1 km with a sampling interval of 0.2 km. We conduct the inversion with three frequency bands—{2.0, 2.5, 3.0} Hz, {5.0, 6.0, 7.0} Hz, and {7.0, 8.0, 9.0} Hz. We discretize the model with 0.05 km grids. The observed data contain 3% Gaussian noise. We use the same optimization strategy as that is used in the Marmousi example.

Fig. 5(b) shows the initial model for FWI and WRI. We use the same strategy as the previous example to initialize LRWI. Fig. 6 shows the results of FWI, WRI with the best selection of $\beta_1 = 1e-4$, and LRWI with the best selection of $\beta_1 = 1e-8$ and $\beta_2 = 1e-12$. Clearly, FWI and WRI converge to the local minima, whereas LRWI converges to the correct solution.

We also investigate the robustness of the three methods with respect to the starting frequency. We vary the starting frequency from 0.5 to 3.0 Hz and use the initial model in Fig. 5(b). Fig. 6(d) shows the relative model errors versus the starting frequency for all the three methods. Again, the necessary lowest frequency for LRWI (2.5 Hz) is significantly higher than that of FWI (0.5 Hz) and WRI (1.0 Hz).
We presented the promising initial results of LRWI in mitigating problems of local minima, while some aspects of the approach warrant further investigations.

1) While our analysis provides unit-free scales for $\lambda$ and $\gamma$ and results imply that selecting $\lambda$ and $\gamma$ to be a small fraction of the scales at the early stage yields plausible results, a more solid justification of this observation is desirable for further applications.

2) While LRWI can conduct inversion with low-frequency data and produce good initial models for FWI with a small computational cost, applications of LRWI to high-frequency data can be valuable when low-frequency data are not available. Therefore, a fast solver for (12) would be worthwhile.

3) Finally, if the storage and computational cost are not bottlenecks, the rank of the lifted matrices could be potentially worth exploring.

V. Conclusion
We presented a “Lift” and “Relax” approach for waveform inversion problems. LRWI is based on a PDE relaxation and a rank-2 variable relaxation, which results in an unconstrained optimization problem with respect to a rank-2 matrix that contains both lifted model parameters and wavefields. We used the variable projection method to develop an efficient optimization solver for the problem. Numerical examples showed that with an acceptable additional computational cost, LRWI can conduct successful inversions with higher frequency data and poorer initial models compared with FWI and WRI.

REFERENCES
[1] A. Tarantola and B. Valette, “Generalized nonlinear inverse problems solved using the least squares criterion,” Rev. Geophys., vol. 20, no. 2, pp. 219–232, 1982.
[2] R. G. Pratt, “Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model,” Geophysics, vol. 64, no. 3, pp. 888–901, May 1999.
[3] J. Virieux and S. Operto, “An overview of full-waveform inversion in exploration geophysics,” Geophysics, vol. 74, no. 6, pp. WCC1–WCC26, Nov. 2009, doi: 10.1190/1.3238367.
[4] D. V. Vigh, W. E. S. Starr, and K. D. Kenneth Dingwall, “3D prestack time domain full waveform inversion,” in Proc. 71st EAGE Conf. Exhib. Incorporating, 2009, p. 217.
[5] M. Warner, T. Nangoo, N. Shah, A. Umpleby, and J. Morgan, “Full-waveform inversion of cycle-skipped seismic data by frequency downshifting,” in Proc. SEG Tech. Program Expanded Abstracts, Sep. 2013, pp. 903–907.
[6] M. Cara and J. J. Lévéque, “Waveform inversion using secondary observables,” Geophys. Lett., vol. 14, no. 10, pp. 1046–1049, Oct. 1987.
[7] T. Van Leeuwen and W. A. Mulder, “A correlation-based misfit criterion for wave-equation traveltime tomography,” Geophys. J. Int., vol. 182, no. 3, pp. 1383–1394, Sep. 2010.
[8] R.-S. Wu, J. Luo, and B. Wu, “Ultra-low-frequency information in seismic data and envelope inversion,” in Proc. SEG Tech. Program Expanded Abstracts, Sep. 2013, pp. 3078–3082.
[9] B. Engquist and B. D. Froese, “Application of the wasserstein metric to seismic signals,” Commun. Math. Sci., vol. 12, no. 5, pp. 979–988, 2014.
[10] M. Warner and L. Guasch, “Adaptive waveform inversion: Theory,” Geophysics, vol. 81, no. 6, pp. R439–R445, Nov. 2016.
[11] Y. Yang, B. Engquist, J. Sun, and B. F. Hamfeldt, “Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion,” Geophysics, vol. 83, no. 1, pp. R43–R62, 2018.
[12] T. V. Leeuwen and F. J. Herrmann, “A penalty method for PDE-constrained optimization in inverse problems,” Inverse Problems, vol. 32, no. 1, Jan. 2016, Art. no. 015007.
[13] G. Huang, R. Nammour, and W. Symes, “Full-waveform inversion via source-receiver extension,” Geophysics, vol. 82, no. 3, pp. R153–R171, May 2017.
[14] Z. Fang, R. Wang, and F. J. Herrmann, “Source estimation for wavefield-reconstruction inversion,” Geophysics, vol. 83, no. 4, pp. R345–R359, Jul. 2018.
[15] Z. Fang, C. Da Silva, R. Kuske, and F. J. Herrmann, “Uncertainty quantification for inverse problems with weak partial-differential-equation constraints,” Geophysics, vol. 83, no. 6, pp. R629–R647, 2018.
[16] W. W. Symes, “Migration velocity analysis and waveform inversion,” Geophys. Prospecting, vol. 56, no. 6, pp. 765–790, 2008.
[17] Y. Li, B. Biondi, R. Clapp, and D. Nichols, “Wave-equation migration velocity analysis for vti models,” Geophysics, vol. 79, no. 3, pp. WA59–WA68, 2014.
[18] A. Cosse, S. D. Shank, and L. Demanet, “A short note on rank-2 relaxation for waveform inversion,” in Proc. SEG Tech. Program Expanded Abstracts, Aug. 2015, pp. 1344–1350.
[19] J. B. Lasserre, “Global optimization with polynomials and the problem of moments,” SIAM J. Optim., vol. 11, no. 3, pp. 796–817, Jan. 2001.
[20] M. Laurent, “Sums of squares, moment matrices and optimization over polynomials,” in Emerging Applications of Algebraic Geometry, New York, NY, USA: Springer-Verlag, 2009, pp. 157–270.
[21] G. Golub and V. Pereyra, “Separable nonlinear least squares: The variable projection method and its applications,” Inverse Problems, vol. 19, no. 2, p. R1, 2003.
[22] J. Nocedal and S. J. Wright, Numerical Optimization. New York, NY, USA: Springer-Verlag, 2006.
[23] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” Geophysics, vol. 60, no. 5, pp. 1475–1473, 1995.

Zhilong Fang received the B.Sc. and M.Sc. degrees in mathematics from Tsinghua University, Beijing, China, in 2010 and 2012, respectively, and the Ph.D. degree in geophysics from the University of British Columbia, Vancouver, BC, Canada, in 2018. He is a Post-Doctoral Associate with the Department of Mathematics and Earth Resources Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. His research interests are in inverse problems, machine learning, uncertainty quantification, seismic imaging, and signal processing.

Laurent Demanet is a Professor with the Department of Mathematics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He holds a joint appointment with the Department of Earth, Atmospheric, and Planetary Sciences, where he is the Director of MIT’s Earth Resources Laboratory. His research interests include applied analysis, scientific computing, machine learning, inverse problems, and wave propagation.

Dr. Demanet is a recipient of a Sloan research fellowship, a CAREER award from NSF, and a Young Investigator award from Air Force Office of Scientific Research (AFOSR).