A numerical solver for least-squares sub-problems in 3D wavefield reconstruction inversion and related problem formulations.

Bas Peters*, University of British Columbia, Vancouver, Canada and Felix J. Herrmann, Georgia Institute of Technology

SUMMARY

Recent years saw a surge of interest in seismic waveform inversion approaches based on quadratic-penalty or augmented-Lagrangian methods, including Wavefield Reconstruction Inversion. These methods typically need to solve a least-squares sub-problem that contains a discretization of the Helmholtz equation. Memory requirements for direct solvers are often prohibitively large in three dimensions, and this limited the examples in the literature to two dimensions. We present an algorithm that uses iterative Helmholtz solvers as a black-box to solve the least-squares problem corresponding to 3D grids. This algorithm enables Wavefield Reconstruction Inversion and related formulations, in three dimensions. Our new algorithm also includes a root-finding method to convert a penalty into a constraint on the data-misfit without additional computational cost, by reusing precomputed quantities. Numerical experiments show that the cost of parallel communication and other computations are small compared to the main cost of solving one Helmholtz problem per source and one per receiver.

INTRODUCTION

Seismic full-waveform inversion (FWI) (Tarantola and Valette, 1982; Pratt et al., 1998; Knibbe et al., 2014) has the potential to provide high-resolution parameter estimates in terms of acoustic velocity and other material properties. However, because of limited source / receiver coverage, and limited frequency bandwidth in the data, it is extremely challenging to estimate geologically realistic and reliable models. As a result, many authors proposed different FWI formulations based on quadratic-penalty and augmented-Lagrangian methods (van Leeuwen and Herrmann, 2013; van Leeuwen et al., 2014; Peters et al., 2014; Peters and Herrmann, 2014; van Leeuwen and Herrmann, 2016; Wang et al., 2016; Esser et al., 2016; Wang and Herrmann, 2017; Fu and Symes, 2017; da Silva and Yao, 2018; Esser et al., 2018; Fang et al., 2018; Aghamiry et al., 2019). Empirical results suggest such formulations are more likely to recover good model estimates under certain conditions.

So far, all presented results for penalty and augmented-Lagrangian based methods are in 2D, likely because the more difficult sub-problems that involve discretizations of wave equations. Whereas time-harmonic FWI has two sub-problems that involve a Helmholtz solve for the forward and adjoint wavefield, penalty-based methods usually need to solve a linear least-squares problem, or corresponding normal equations that contain a Helmholtz block. In 2D, this is not a big issue, as there are fast implementations of direct QR or Cholesky factorizations available in most programming languages. In 3D, however, direct factorization is often not an option due to the

increased memory requirements.

Iterative algorithms for linear systems require less memory. Various specialized solvers and preconditioners for Helmholtz problems are available, but they cannot be applied to a least-squares problem. In this work, we develop an algorithm that solves the least-squares sub-problems that arise in Wavefield Reconstruction Inversion (WRI) and variants. The proposed algorithm can use standardly available Helmholtz solvers for its sub-problems, the other computational work is negligible in terms of computing complexity and time. The new algorithm opens the door to WRI in 3D. We provide timings and a numerical example. Finally, we show the relation between sub-problems of the data-constrained and the penalty formulation. This allows us to compute the correspondence explicitly at low computational cost. Our algorithm is, therefore, suitable to solve sub-problems of certain versions of both the quadratic penalty and data-constrained formulation of seismic inversion.

WAVEFIELD RECONSTRUCTION INVERSION AND RELATED FORMULATIONS

Many partial-differential-equation (PDE) constrained problems often (Haber et al., 2000; Biros and Ghattas, 2005; Epanomeritakis et al., 2008) use the following formulation

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|\mathbf{Pu} - \mathbf{d}\|_2^2 \quad \text{subject to} \quad \mathbf{H}(\mathbf{m})\mathbf{u} = \mathbf{q}. \qquad (1)$$

Here, we stated the problem for one frequency and one source. This problem represents the minimization of the misfit between observed data $\mathbf{d} \in \mathbb{C}^M$ and the predicted data $\mathbf{P}u$. The matrix $\mathbf{P} \in \mathbb{R}^{M \times N}$ selects from the wavefield $\mathbf{u} \in \mathbb{C}^N$ on N grid-points, the field values at the M receiver locations. The equality constraints enforce that the Helmholtz equation holds in discrete form, where $\mathbf{q} \in \mathbb{C}^N$ is the source vector and $\mathbf{H}(\mathbf{m}) \in \mathbb{C}^{N \times N}$ is the Helmholtz discretization that depends on the unknown model parameters (velocity) $\mathbf{m} \in \mathbb{R}^N$.

FWI is mostly formulated using a reduced Lagrangian, or, reduced-space formulation, where the constraints are always satisfied as part of the unconstrained problem

$$\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{PH}(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}\|_2^2. \qquad (2)$$

van Leeuwen and Herrmann (2013, 2016) propose algorithms based on the quadratic penalty form

$$\phi_\lambda(\mathbf{m},\mathbf{u}) = \frac{1}{2}\|\mathbf{H}(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{Pu} - \mathbf{d}\|_2^2 \qquad (3)$$

with scalar penalty parameter $\lambda > 0$. From here, there are many ways to proceed, including directly solving (3), modified gradient (Kleinman and den Berg, 1992) and contrast source (van den Berg and Kleinman, 1997), alternating minimization

van Leeuwen and Herrmann (2013), and variable projected version (van Leeuwen and Herrmann, 2016) that solves a reduced version of (3) by projecting out the field variables **u**, see (Golub and Pereyra, 1973; Aravkin and van Leeuwen, 2012; Aravkin et al., 2018). This last approach minimizes

$$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\|\mathbf{H}(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{P}\bar{\mathbf{u}} - \mathbf{d}\|_2^2, \qquad (4)$$

where $\bar{\mathbf{u}}$ is the solution to

$$\nabla_{\mathbf{u}}\phi_\lambda(\mathbf{u}, \mathbf{m}) = \mathbf{H}(\mathbf{m})^*(\mathbf{H}(\mathbf{m})\mathbf{u} - \mathbf{q}) + \lambda^2\mathbf{P}^*(\mathbf{P}\mathbf{u} - \mathbf{d}) = 0. \qquad (5)$$

The field variables are projected out at every iteration of minimizing (4) over **m**, and this sub-problem is equivalent to the linear least-squares problem

$$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \mathbf{H}(\mathbf{m}) \\ \lambda\mathbf{P} \end{pmatrix} \mathbf{u} - \begin{pmatrix} \mathbf{q} \\ \lambda\mathbf{d} \end{pmatrix} \right\|_2^2. \qquad (6)$$

If we instead use alternating minimization for problem (3), we obtain the same sub-problem for **u** as (6). If we have accurate knowledge about the data noise, it is appealing to minimize the PDE error subject to a data-fit constraint,

$$\min_{\mathbf{m}, \mathbf{u}} \|\mathbf{H}(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 \leq \gamma, \qquad (7)$$

see, e.g., Wang and Herrmann (2017); Fu and Symes (2017). The scalar $\gamma > 0$ depends on the noise level in the observed data. Note that alternating minimization for example, also leads to the same subproblem, because

$$\min_{\mathbf{u}} \|\mathbf{H}\mathbf{u} - \mathbf{q}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 \leq \gamma. \qquad (8)$$

is equivalent to the linear least-squares problem (6) for a certain γ-λ pair (Gander, 1980; Björck, 1996). Moreover, there are algorithms to retrieve the λ to a corresponding γ, see Björck (1996) Sect. 5.3.2.

In conclusion, we see that the least-squares problem (6) or equivalent normal equations occur in multiple problem formulations and solution approaches. In the next few sections, we propose algorithms for solving this sub-problem.

SOLUTION OF THE LEAST-SQUARES SUB-PROBLEM

We do not investigate which problem formulations is the 'best'. This work is about solving the sub-problem for problems discretized on 3D grids. The preceding section is to motivate the problem, and show that this work applies to multiple formulations of waveform inversion.

Our algorithm requires two assumptions that are not restrictive for geophysical applications. *1)* The PDE discretization is full rank so $\mathbf{H}^{-1}$ exists. *2)* there are no two receivers at the same spatial location. This implies the rows of **P** are linearly independent. We also assume that the number of receivers does not exceed a few hundred for memory considerations. We start by rewriting the normal equations from (5) as

$$(\mathbf{I}_N + \lambda^2\mathbf{H}^{-*}\mathbf{P}^*\mathbf{P}\mathbf{H}^{-1})\mathbf{y} = \mathbf{q} + \lambda^2\mathbf{H}^{-*}\mathbf{P}^*\mathbf{d}, \text{ with } \mathbf{H}\mathbf{u} = \mathbf{y}, \quad (9)$$

where $\mathbf{y} \in \mathbb{C}^N$ is a temporary vector and $\mathbf{I}_N$ is the $N \times N$ identity matrix. The complex-conjugate transpose of a matrix is

denoted as $\mathbf{H}^{-*}$. We dropped the dependency of the PDE on **m** because we are interested in solving sub-problems w.r.t. **u** and for a fixed **m**. To better see structure in the equations and make notation more compact, we define

$$\mathbf{W} \equiv \mathbf{H}^{-*}\mathbf{P}^* \in \mathbb{C}^{N \times M}, \qquad (10)$$

which is a tall and dense matrix with a size of the number of grid points $\times$ the number of receivers. Plugging this definition in the preceding rewritten problem statement, we arrive at the identity + low-rank factorized form:

$$(\mathbf{I}_N + \lambda^2\mathbf{W}\mathbf{W}^*)\mathbf{y} = \mathbf{q} + \lambda^2\mathbf{W}\mathbf{d}, \quad \text{with} \quad \mathbf{H}\mathbf{u} = \mathbf{y}. \quad (11)$$

The inverse of this system matrix is known in closed form as the Sherman-Morrison-Woodbury (SMW) identity:

$$(\mathbf{I}_N + \lambda^2\mathbf{W}\mathbf{W}^*)^{-1} = \mathbf{I}_N - \lambda^2\mathbf{W}(\mathbf{I}_M + \lambda^2\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*. \quad (12)$$

Via the SMW identity we rewrite (11) to obtain our final problem formulation as

$$\mathbf{y} = (\mathbf{I}_N - \lambda^2\mathbf{W}(\mathbf{I}_M + \lambda^2\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*)(\mathbf{q} + \lambda^2\mathbf{W}\mathbf{d}) \text{ with } \mathbf{H}\mathbf{u} = \mathbf{y}. \quad (13)$$

The computationally difficult piece is the inverse matrix $(\mathbf{I}_M + \lambda^2\mathbf{W}^*\mathbf{W})^{-1} \in \mathbb{C}^{M \times M}$. While this small matrix has a size of the number of receivers squared, we need access to **W** to compute it. We propose to make a one-time investment and compute **W** column by column: a computational cost of one PDE solve per receiver. This computation does not need to be repeated for every source location. Each of these columns may be computed independently in parallel using iterative (or direct) solvers. These last few observations suggest Algorithm 1 to solve the slightly overdetermined least-squares problem (6). Algorithm 1 computes the fields for all sources and all

---

Algorithm 1 Algorithm for solving problem (6).

---

input:
$\mathbf{H} \in \mathbb{C}^{N \times N}, \mathbf{q} \in \mathbb{C}^N, \mathbf{P} \in \mathbb{R}^{M \times N}, \mathbf{d} \in \mathbb{C}^M$
for each receiver (j) (row in **P**) in parallel **do**
  $\mathbf{H}^*\mathbf{w}_j = \mathbf{p}_j^*$ {solve 1 PDE}
end for
$\mathbf{W} = [\mathbf{w}_1 \, \mathbf{w}_2 \ldots \mathbf{w}_m]$ {distributed matrix}
$\mathbf{S} = (\mathbf{I}_M + \lambda^2\mathbf{W}^*\mathbf{W})^{-1}$
{adjust λ using Algorithm 2 (optional)}
for source (i) in parallel **do**
  $\mathbf{y}_i = (\mathbf{I}_N - \lambda^2\mathbf{W}\mathbf{S}\mathbf{W}^*)(\mathbf{q}_i + \lambda^2\mathbf{W}\mathbf{d}_i)$
  $\mathbf{H}\mathbf{u}_i = \mathbf{y}_i$ {solve 1 PDE}
end for
output: $\mathbf{u}_i$ {fields for all sources}

---

receivers at a cost of 1 PDE per source + 1 PDE solve per receiver. This is different from the 2 PDEs per source location for regular FWI via the adjoint-state method. Therefore, the proposed algorithm is suitable for a moderate number of receivers. To keep the number of sources/receivers low, we may employ simultaneous receivers (Habashy et al., 2011; Peters et al., 2016), similar to simultaneous sources.

A possible problem in Algorithm 1 is the conditioning of $(\mathbf{I}_M + \lambda^2\mathbf{W}^*\mathbf{W})$. Numerical experiments (not shown) suggest that the condition number is often lower than 10^2 for 3D problems on regular grids and values of λ that correspond to a γ of about 10% relative data-fit error in (8).

THE λ-γ RELATION

The relation between the linear least-squares problem (6) and the least-squares problem with quadratic inequality constraints (8) was studied extensively (Gander, 1980; Björck, 1996), also algorithmically. Our assumptions stated earlier suffice to guarantee that there is a λ-γ pair that causes the solutions of the two problems to be the same (Björck, 1996, Thm. 5.3.1). Assumption 2 implies $\min_{\mathbf{u}} \|P\mathbf{u} - \mathbf{d}\| = 0 \leq \gamma$, so at least one feasible solutions exist. To reveal the connection between γ and λ, we turn to the Lagrangian corresponding to (8),

$$\mathscr{L}(\mathbf{u}) = \|\mathbf{Hu} - \mathbf{q}\|_2^2 + \mu(\|\mathbf{Pu} - \mathbf{d}\|_2^2 - \gamma^2), \quad (14)$$

with scalar Lagrangian multiplier μ. A necessary condition for optimality is $\nabla_{\mathbf{u}}\mathscr{L} = 0$, i.e.,

$$(\mathbf{H}^*\mathbf{H} + \mu\mathbf{P}^*\mathbf{P})\mathbf{u}_\mu = \mathbf{H}^*\mathbf{q} + \mu\mathbf{P}^*\mathbf{d}, \quad (15)$$

where $\mathbf{u}_\mu$ satisfies the constraint $\|\mathbf{Pu}_\mu - \mathbf{d}\|_2 = \gamma$. We can find $\mathbf{u}_\mu$ at a root (zero) of the scalar equation (the secular equation)

$$f(\mu) = \|\mathbf{Pu}_\mu - \mathbf{d}\|_2 - \gamma, \quad (16)$$

where we require that $\mathbf{u}_\mu$ is a solution of (15). Secant root-finding needs a solution of (15) at every iteration; very costly without factorizations. The penalty-constraint relation and root-finding procedure are well-established techniques. Our contribution is that we derive a way to perform the rootfinding without much additional computational cost and avoid re-solving the normal equations (15) at every secant iteration.

We need to find a root of the secular equation (16) which depends on $\|\mathbf{Pu}_\mu - \mathbf{d}\|_2$. However, we do not need to know the full solution $\mathbf{u}_\mu$; the restriction of the solution at the receivers is sufficient to compute $\mathbf{Pu}_\mu$. After some manipulations, we rewrite the normal equations (15) as

$$\mathbf{Pu}_\mu - \mathbf{d} = \mathbf{S}(\mathbf{W}^*\mathbf{q} - \mathbf{d}). \quad (17)$$

We already precomputed $\mathbf{W}$ and $\mathbf{W}^*\mathbf{W}$, so we can adjust λ and recompute $\mathbf{S} = (\mathbf{I}_M + \lambda^2\mathbf{W}^*\mathbf{W})^{-1}$ cheaply. All operations in the secant method are on small matrices at a computational cost insignificant compared to solving PDEs. The corresponding Algorithm 2 fits in the middle of Algorithm 1. The output λ is such that the final data-misfit satisfies $\|\mathbf{Pu} - \mathbf{d}\|_2 \leq \gamma$.

COMPUTATIONS, MEMORY & COMMUNICATION

The two PDE solves, one per receiver and one per source, are the main computational cost. Precomputations involving the receivers are a one-time investment (for fixed $\mathbf{m}$ and one frequency). Other computations related to the SMW identity are not significant. The root-finding procedure, which is typically very expensive computationally, is easy to compute because of the precomputed quantities.

The primary memory requirement is the storage of $\mathbf{W}$ of size (number of grid points) $\times$ (number of receivers). This matrix may be distributed over the different compute nodes, however. There is only one such matrix for all sources and receivers and storage is typically not a problem if we have up to a few hundred receivers and at least a couple of compute nodes.

The parallel communication occurs when computing matrix-vector products with $\mathbf{W}$ and $\mathbf{W}^*$ because these matrices are distributed over multiple compute nodes. In the numerical examples, we show that the communication time is not large compared to the time it takes to solve PDEs.

Algorithm 2 Algorithm (secant) to obtain the penalty parameter λ corresponding to data fit constraint parameter γ using precomputed quantities.

Input: $\mathbf{W}$, $\mathbf{X} = \mathbf{W}^*\mathbf{W}$, γ, $\mathbf{v} \equiv \mathbf{W}^*\mathbf{q} - \mathbf{d}$
$f(\lambda) \equiv \|\mathbf{S}(\lambda)\mathbf{v}\|_2 - \gamma$, $\mathbf{S}(\lambda) = (\mathbf{I}_M + \lambda_k^2\mathbf{X})^{-1}$
$\lambda_0 = 0$, $\lambda_1 \in (0, \lambda_*)$, $k = 1$
$\mathbf{S}_0(\lambda) = (\mathbf{I}_M + \lambda_0^2\mathbf{X})^{-1}$,
while $\|\mathbf{S}_k(\lambda)\mathbf{v}\|_2 > \gamma$ **do**
$\quad \lambda_{k+1} = \lambda_k - \frac{f(\lambda_k)(\lambda_k - \lambda_{k-1})}{f(\lambda_k) - f(\lambda_{k-1})}$
$\quad k \leftarrow k + 1$
end while
Output: λ

NUMERICAL EXAMPLES

The Helmholtz solves are considered a black-box in this work so that we can use any solver, in principle. Our examples use the CGMN algorithm (Björck and Elfving, 1979; Gordon and Gordon, 2008), which is equivalent to a preconditioned conjugate-gradient method where a double Kaczmarz sweep acts as the matrix-vector product. This method has been used to solve serveral wave-propagation problems and discretizations (Gordon and Gordon, 2012, 2013; Turkel et al., 2013; van Leeuwen and Herrmann, 2014; Gordon et al., 2015; Li et al., 2015). The numerical experiments are in complex double precision and were carried out on nodes with two CPUs of ten cores (Intel Ivy Bridge 2.8 GHz E5-2680v2) and 128 GB of memory per node. Parallel communication was handled by Matlab Parallel Computing Toolbox. PDE solves are in C++ (Silva and Herrmann, 2019) using a relative residual of $\varepsilon = 10^{-4}$ as stopping condition.

Frequency & grid scaling. Figure 1 shows the time as a function of frequency between 2 and 16 Hz. We discretize the 6 km^3 model such that we always use six gridpoints per shortest spatial wavelength. 'comp U' and 'comp W' indicate the time to solve the two sets of PDEs. The figure shows that parallel communication and other computations take a small amount of time ('other' in figure). The computations use 8 compute nodes and there are 64 sources and 64 receivers. The timings are thus for 2×64 PDEs, 8 per compute node.

Weak parallel scaling with sources & receivers. The parallel communication time depends on the number of grid points, which increases with frequency, and the number of receivers. The previous experiment showed a short communication time in case of 64 receivers. Now, we fix the frequency (and grid) to 8 Hertz and increase the number of sources & receivers as well as the number of compute nodes. We keep solving 8 PDEs per node. Figure 2 shows timings when we increase the number of sources and receivers from 8 to 128 while increasing the number of compute nodes from one to 16. As expected, the time for things other than PDE solves starts increasing, but is
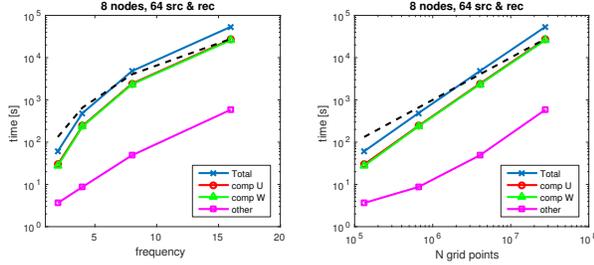
Figure 1: Time-frequency scaling for fixed number of sources, receivers, and resources.

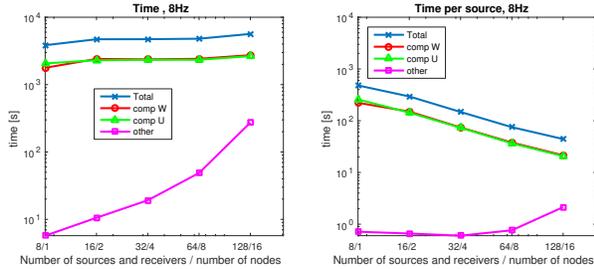small for the assumed number of sources and receivers.

Figure 2: Weak parallel scaling for an increasing number of sources and receivers, while also increasing the number of compute nodes.

3D WRI A simple 3D WRI example shows that a fixed relative residual for the inexact Helmholtz sub-problems is sufficient for Algorithm 1 to work properly. The true and initial model are shown in Figures 3a and 3b.

The synthetic observed data were generated on a fine grid using a relative residual a factor $\times 100$ more accurate than we use for the inverse problem. There is data from $2 - 5$ Hertz, and we use a grid suitable for these frequencies, which is coarser than the grid for data generation. There are 16 sources in two corners and 16 receivers in the other two corners. We formulate the inverse problem as WRI in (4). The optimization uses bound constraints implemented via the spectral projected gradient algorithm (SPG) (Birgin et al., 1999). The result in Figure 4 recovers the true model relatively well, except for some artifacts in the areas between the receiver boreholes and in between the source boreholes. These areas are difficult to image because very little wave energy propagates through them from source to receiver. Some regularization may be required to improve the result.

CONCLUSIONS

We constructed a three-step algorithm for solving least-squares sub-problem in penalty-based formulations for 3D seismic waveform inversion. The algorithm reduces the least-squares problem to an identity+low-rank factorization. The main computational ingredient is any iterative Helmholtz solver to solve one PDE per source plus one per receiver. We also show that our algorithm can solve sub-problems originating from both penalty
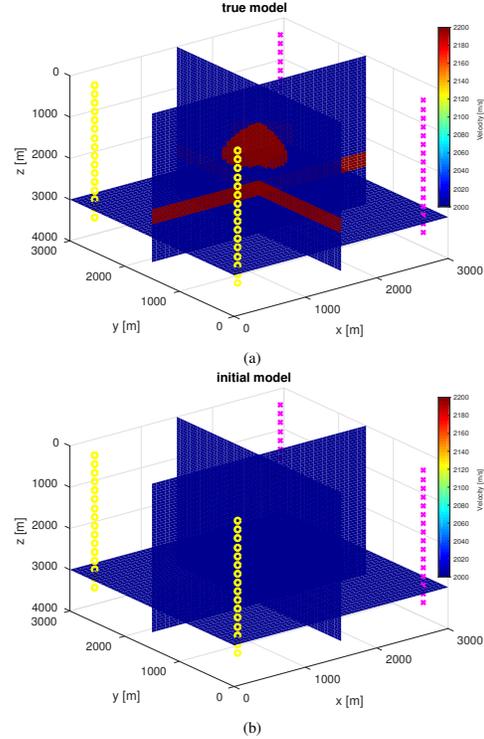

(a)

(b)

Figure 3: True model and initial guess for the 3D acoustic parameter estimation example. Yellow circles represent sources, purple crosses are receivers.

and data-constrained formulations of waveform inversion. A root finding algorithm enables this functionality and requires little additional computational cost because it employs quantities that are by-products of our algorithm. Timings, scaling, and a 3D wavefield reconstruction inversion example shows that the proposed algorithm spends only a small amount of time on parallel communication and computations other than PDE solves.
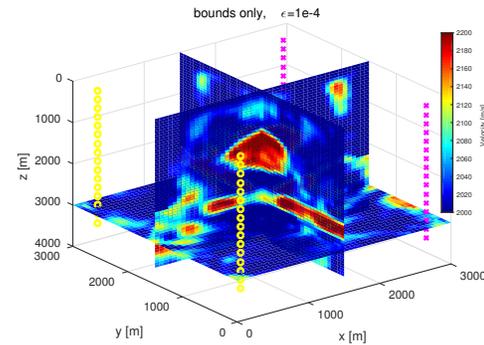

Figure 4: Estimated model with bound constraints and using fixed accuracy (ε) for the Helmholtz sub-problems.

REFERENCES

Aghamiry, H. S., A. Gholami, and S. Operto, 2019, Improving full waveform inversion by wavefield reconstruction with the alternating direction method of multipliers: GEOPHYSICS, **0**, 1–79.

Aravkin, A. Y., D. Drusvyatskiy, and T. van Leeuwen, 2018, Efficient quadratic penalization through the partial minimization technique: IEEE Transactions on Automatic Control, **63**, 2131–2138.

Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: Inverse Problems, **28**, 115016.

Birgin, E. G., J. M. Martínez, and M. Raydan, 1999, Nonmonotone spectral projected gradient methods on convex sets: SIAM J. on Optimization, **10**, 1196–1211.

Biros, G., and O. Ghattas, 2005, Parallel Lagrange–Newton–Krylov–Schur Methods for PDE-Constrained Optimization. Part II: The Lagrange–Newton Solver and Its Application to Optimal Control of Steady Viscous Flows: SIAM Journal on Scientific Computing, **27**, 714–739.

Björck, Å., 1996, Numerical methods for least squares problems: Society for Industrial and Applied Mathematics.

Björck, Å., and T. Elfving, 1979, Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations: BIT, **19**, 145–163.

da Silva, N. V., and G. Yao, 2018, Wavefield reconstruction inversion with a multiplicative cost function: Inverse Problems, **34**, 015004.

Epanomeritakis, I., V. Akçelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: Inverse Problems, **24**, 034015.

Esser, E., L. Guasch, F. J. Herrmann, and M. Warner, 2016, Constrained waveform inversion for automatic salt flooding: The Leading Edge, **35**, 235–239.

Esser, E., L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, 2018, Total variation regularization strategies in full-waveform inversion: SIAM Journal on Imaging Sciences, **11**, 376–406.

Fang, Z., C. D. Silva, R. Kuske, and F. J. Herrmann, 2018, Uncertainty quantification for inverse problems with weak partial-differential-equation constraints: GEOPHYSICS, **83**, R629–R647.

Fu, L., and W. W. Symes, 2017, A discrepancy-based penalty method for extended waveform inversion: Geophysics, **82**, R287–R298.

Gander, W., 1980, Least squares with a quadratic constraint: Numerische Mathematik, **36**, 291–307.

Golub, G., and V. Pereyra, 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM Journal on Numerical Analysis, **10**, 413–432.

Gordon, D., and R. Gordon, 2008, Cgmn revisited: Robust and efficient solution of stiff linear systems derived from elliptic partial differential equations: ACM Trans. Math. Softw., **35**, 18:1–18:27.

———, 2012, Parallel solution of high frequency helmholtz equations using high order finite difference schemes: Applied Mathematics and Computation, **218**, 10737 – 10754.

———, 2013, Robust and highly scalable parallel solution of the helmholtz equation with large wave numbers: Journal of Computational and Applied Mathematics, **237**, 182 – 196.

Gordon, D., R. Gordon, and E. Turkel, 2015, Compact high order schemes with gradient-direction derivatives for absorbing boundary conditions: Journal of Computational Physics, **297**, 295 – 315.

Habashy, T. M., A. Abubakar, G. Pan, and A. Belani, 2011, Source-receiver compression scheme for full-waveform seismic inversion: GEOPHYSICS, **76**, R95–R108.

Haber, E., U. M. Ascher, and D. Oldenburg, 2000, On optimization techniques for solving nonlinear inverse problems: Inverse Problems, **16**, 1263–1280.

Kleinman, R., and P. den Berg, 1992, A modified gradient method for two- dimensional problems in tomography: Journal of Computational and Applied Mathematics, **42**, 17 – 35.

Knibbe, H., W. Mulder, C. Oosterlee, and C. Vuik, 2014, Closing the performance gap between an iterative frequency-domain solver and an explicit time-domain scheme for 3d migration on parallel architectures: Geophysics, **79**, S47–S61.

Li, Y., L. Métivier, R. Brossier, B. Han, and J. Virieux, 2015, 2d and 3d frequency-domain elastic wave modeling in complex media with a parallel iterative solver: Geophysics, **80**, T101–T118.

Peters, B., F. Herrmann, and T. V. Leeuwen, 2016, Parallel reformulation of the sequential adjoint-state method: SEG Technical Program Expanded Abstracts 2016, 1411–1415.

Peters, B., and F. J. Herrmann, 2014, A sparse reduced hessian approximation for multi-parameter wavefield reconstruction inversion, *in* SEG Technical Program Expanded Abstracts 2014: Society of Exploration Geophysicists, 1206–1210.

Peters, B., F. J. Herrmann, and T. van Leeuwen, 2014, Wave-equation based inversion with the penalty method-adjoint-state versus wavefield-reconstruction inversion: Presented at the 76th EAGE Conference and Exhibition 2014.

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

Silva, C. D., and F. Herrmann, 2019, A unified 2d/3d large-scale software environment for nonlinear inverse problems: ACM Trans. Math. Softw., **45**, 7:1–7:35.

Tarantola, A., and A. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of

Geophysics and Space Physics, **20**, 129–232.

Turkel, E., D. Gordon, R. Gordon, and S. Tsynkov, 2013, Compact 2d and 3d sixth order schemes for the helmholtz equation with variable wave number: Journal of Computational Physics, **232**, 272 – 287.

van den Berg, P. M., and R. E. Kleinman, 1997, A contrast source inversion method: Inverse Problems, **13**, 1607–1620.

van Leeuwen, T., F. Herrmann, and B. Peters, 2014, A new take on fwi-wavefield reconstruction inversion: Presented at the 76th EAGE Conference and Exhibition 2014.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

——, 2014, 3d frequency-domain seismic inversion with controlled sloppiness: SIAM Journal on Scientific Computing, **36**, S192–S217.

——, 2016, A penalty method for pde-constrained optimization in inverse problems: Inverse Problems, **32**, 015007.

Wang, C., D. Yingst, P. Farmer, and J. Leveille, 2016, Full-waveform inversion with the reconstructed wavefield method: SEG Technical Program Expanded Abstracts 2016, 1237–1241.

Wang, R., and F. Herrmann, 2017, A denoising formulation of full-waveform inversion, *in* SEG Technical Program Expanded Abstracts 2017: Society of Exploration Geophysicists, 1594–1598.