Application of misfit-based model space coordinate system design to seismic full waveform inversion

Kris Innanen

ABSTRACT

The re-parameterizations we encounter regularly in seismic full waveform inversion can all be formally identified as coordinate transforms between suitably general coordinate systems. For instance, changes-of-variable between elastic/acoustic parameter classes, and the variable changes underlying the model space pre-conditioning methods of Harlan, Claerbout and Fomel, can both be shown to be transforms between oblique-rectilinear coordinate systems when covariant notation is employed. With transformation rules in place for updates and displacements (contravariant vectors), and sensitivities and gradients (covariant vectors), as well as objective functions (scalars) and Hessian operators (2nd rank covariant tensors), we may then seek new and potentially useful re-parameterizations. Coordinate systems within which the Hessian is close to the identity have vastly improved optimization properties. Using a newly-derived procedure for designing transformation matrices mapping to such systems, we provide an early examination of re-parameterized FWI.

INTRODUCTION

In this paper the purpose will be to discuss re-parameterization and optimization of seismic full waveform inversion. Parameterization has long been recognized to dramatically alter the numerical convergence properties of seismic FWI (Anagaw and Sacchi, 2018), and the ability to move from one parameter class to another can enhance interpretability of the results (Hu and Innanen, 2020). Furthermore regularized objective functions, which penalize certain types of model structure or behaviour, can be much more effectively incorporated by suitable model space preconditioning, which is in effect a change in parameterization as well. In this report Keating and Innanen (2020) discusses simultaneous updating of source properties and medium properties, and it emerges that model unknowns defined grid point by grid point produce significant nonlinearities in the problem; re-parameterization in terms of basis functions with larger spatial extent is needed.

That re-parameterizations, or changes-of-variable, in seismic inversion, are coordinate transforms is more or less immediately clear in an imprecise sense. What is more interesting is that they are coordinate transforms in a mathematically precise sense. It is necessary to allow the coordinate systems linked to one another to be non-Cartesian in order to establish this, but upon doing so we find this relatively straightforwardly. Part of this paper is involved with establishing this fact. We do so by analytically formulating a discrete frequency domain FWI scheme, keeping it unrealistically small in order that it be treatable element by element. When we change this system from one parameterization to another, it immediately emerges that model updates (i.e., displacements) transform as contravariant vectors, and gradients transform as covariant vectors. We also establish that model space pre-conditioning mathematics emerges naturally when we assume that the regularization operator mediates a transformation between oblique-rectilinear coordinate systems.

With this set out, it becomes possible to apply the coordinate transform design program to seismic FWI, seeking one of many possible coordinate transforms on model space under which the Hessian is a Kronecker delta function, meaning that relatively cross-talk free updates should be possible via steepest descent optimization (in the new system). We validate the idea with a small scalar acoustic FWI example, and, in steepest descent optimization carried out in the transform domain, observe the appearance of illumination compensation not unlike that occurring in full Gauss-Newton updates.

FREQUENCY DOMAIN FULL WAVEFORM INVERSION

To establish the legitimacy of a covariant formulation of model re-parameterization, we first must set up an FWI problem in such a way that it is exposed to analysis. We do this first with a very small (but mathematically valid) frequency domain simulation environment, and with it building up both scalar objective function, update vector, gradient vector, and Hessian tensor quantities. The approach is essentially that of Pratt (1999).

Simulations

The variable density wave equation is

$$\nabla \cdot \left(\frac{1}{\rho} \nabla \right) u' - \frac{1}{\kappa} \frac{\partial^2}{\partial t^2} u' = f', \tag{1}$$

where u' and f' are the wavefield and source function in the space and time domain. The 2D version of this equation in the frequency domain is

$$[\partial_x, \partial_z] \cdot \left(s_\rho \left[\begin{array}{c} \partial_x u \\ \partial_z u \end{array} \right] \right) + \omega^2 s_\kappa u = f, \tag{2}$$

where $s_\rho = {}^1/\rho$, $s_\kappa = {}^1/\kappa$, and u and f are the temporal Fourier transforms of u' and f' respectively. This can be formally expressed as an operator equation

$$\mathcal{A}(s)\mathbf{u} = \mathbf{f}, \tag{3}$$

with the solution $\mathbf{u} = \mathcal{A}^{-1}(s)\mathbf{f}$.

Structure of the discrete forward modelling problem

Let us take a definite, though unrealistically small, example of a simulation based on the equations above. We discretize a 2D volume over a 2×2 grid, as illustrated in Figure 1. In the discretization, the medium properties and wave field values are positioned with discrete x_j and z_i values, with i and j ranging over 1 and 2. In the illustration we include an additional single index (in the upper left corner of each grid cell) ranging from 1 to 4. Derivatives of u with respect to x are approximated with the finite difference formulas:

$$\partial_x u = \frac{u(x + \Delta x, z) - u(x, z)}{\Delta x}, \quad \partial_x^2 u = \frac{u(x + \Delta x, z) - 2u(x, z) + u(x - \Delta x, z)}{\Delta x^2}.$$

$$
\begin{array}{cc}
x_1 & x_2
\end{array}
$$

	x_1	x_2
z_1	**1** u, s_ρ, s_κ	**2** u, s_ρ, s_κ
z_2	**3** u, s_ρ, s_κ	**4** u, s_ρ, s_κ
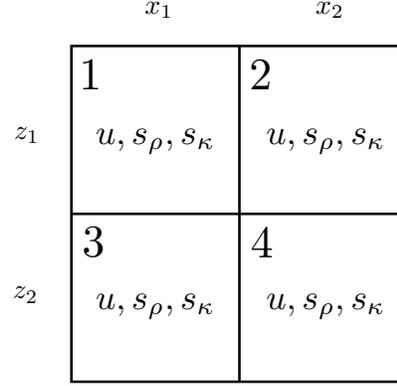
FIG. 1. Prototype 2x2 discretization scheme for frequency domain forward modeling with the variable-density acoustic wave equation.

The z derivatives, and all derivatives of s_ρ, are approximated similarly. The approximations are substituted into equation (2), producing a set of simultaneous equations for the four values of u. Choosing units such that $\Delta x = \Delta z = 1$, these equations are

$$
\left[s_{\rho_{i,j+1}} - s_{\rho_{i,j}} \right] \left[u_{i,j+1} - u_{i,j} \right] + \left[s_{\rho_{i+1,j}} - s_{\rho_{i+1,j}} \right] \left[u_{i+1,j} - u_{i,j} \right] + \omega^2 s_{\kappa_{i,j}} u_{i,j} = f_{i,j}, \quad (4)
$$

with i and j ranging over 1, 2. The four equations are arranged in matrix form using the numbering convention in Figure 1, where rather than labelling a medium property $s_{\rho_{2,1}}$, for example, we just put s_{ρ_3}. Doing this, and setting u values outside the domain to zero, we can contain the four equations of (4) in the system

$$
\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix}, \quad \text{and} \quad \mathcal{A}(s) = \begin{bmatrix} K_1 & s_{\rho_2} & s_{\rho_3} & 0 \\ s_{\rho_2} & K_2 & 0 & s_{\rho_4} \\ s_{\rho_3} & 0 & K_3 & s_{\rho_4} \\ 0 & s_{\rho_4} & s_{\rho_4} & K_4 \end{bmatrix}, \quad (5)
$$

where $K_1 = \omega^2 s_{\kappa_1} - 2s_{\rho_1} - s_{\rho_2} - s_{\rho_3}$, $K_2 = \omega^2 s_{\kappa_2} - 2s_{\rho_2} - s_{\rho_4}$, $K_3 = \omega^2 s_{\kappa_3} - 2s_{\rho_3} - s_{\rho_4}$, and $K_4 = \omega^2 s_{\kappa_4} - 2s_{\rho_4}$. The solution $\mathbf{u} = \mathcal{A}^{-1}(s)\mathbf{f}$ at a given frequency ω is then explicitly

$$
\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} K_1 & s_{\rho_2} & s_{\rho_3} & 0 \\ s_{\rho_2} & K_2 & 0 & s_{\rho_4} \\ s_{\rho_3} & 0 & K_3 & s_{\rho_4} \\ 0 & s_{\rho_4} & s_{\rho_4} & K_4 \end{bmatrix}^{-1} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix}. \quad (6)
$$

Indicial notation

We have organized the four wavefield values into the column vector $\mathbf{u}$, which we can represent as u_i in indicial notation. Let us now do the same with the model values, creating

a vector **s** that is 8 points long (since there are two unknowns at each of 4 locations):

$$\mathbf{s} = s^\mu = \begin{bmatrix} s^1 \\ s^2 \\ s^3 \\ s^4 \\ s^5 \\ s^6 \\ s^7 \\ s^8 \end{bmatrix} = \begin{bmatrix} s_{\rho 1} \\ s_{\rho 2} \\ s_{\rho 3} \\ s_{\rho 4} \\ s_{\kappa 1} \\ s_{\kappa 2} \\ s_{\kappa 3} \\ s_{\kappa 4} \end{bmatrix}. \tag{7}$$

At this stage we have introduced Greek superscript indices, because (unlike **u**) this vector is an element of model space, which we are treating in terms of the more general coordinate geometry. The impedance matrix in equation (5) is then

$$\mathcal{A}(s) = \begin{bmatrix} \omega^2 s^5 - 2s^1 - s^2 - s^3 & s^2 & s^3 & 0 \\ s^2 & \omega^2 s^6 - 2s^2 - s^4 & 0 & s^4 \\ s^3 & 0 & \omega^2 s^7 - 2s^3 - s^4 & s^4 \\ 0 & s^4 & s^4 & \omega^2 s^8 - 2s^4 \end{bmatrix},$$

or, for compactness,

$$\mathcal{A}(s) = a_{ij}(s) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}. \tag{8}$$

The matrix is labelled with Roman indices, since, as with u_i, the row and column dimensions here are not related to model space (though each element is in general a function of the elements of s^μ). In indicial notation (6) can be written

$$u_i = a_{ij}^{-1}(s) f_j. \tag{9}$$

For Roman indices summation is implied but the upper/lower notation is not used.

Velocity-density parameterization

The previous results comprise the simulation in a modulus-density parameterization. To instead express the problem in terms of wave velocity c and density, we can make use of $c = \sqrt{\kappa/\rho}$, and define new coordinates $r_\rho = 1/\rho$ and $r_c = 1/c^2$, which are related by

$$s_\rho = r_\rho, \quad s_\kappa = r_c r_\rho. \tag{10}$$

The re-parameterized problem is then

$$\begin{bmatrix} V_1 & r_{\rho 2} & r_{\rho 3} & 0 \\ r_{\rho 2} & V_2 & 0 & r_{\rho 4} \\ r_{\rho 3} & 0 & V_3 & r_{\rho 4} \\ 0 & r_{\rho 4} & r_{\rho 4} & V_4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix}, \tag{11}$$

where $V_1 = \omega^2 r_{c_1} r_{\rho_1} - 2r_{\rho_1} - r_{\rho_2} - r_{\rho_3}$, $V_2 = \omega^2 r_{c_2} r_{\rho_2} - 2r_{\rho_2} - r_{\rho_4}$, $V_3 = \omega^2 r_{c_3} r_{\rho_3} - 2r_{\rho_3} - r_{\rho_4}$ and $V_4 = \omega^2 r_{c_4} r_{\rho_4} - 2r_{\rho_4}$. Again organizing all 8 model parameters into a single vector

$$\mathbf{r} = r^\mu = \begin{bmatrix} r^1 \\ r^2 \\ r^3 \\ r^4 \\ r^5 \\ r^6 \\ r^7 \\ r^8 \end{bmatrix} = \begin{bmatrix} r_{\rho_1} \\ r_{\rho_2} \\ r_{\rho_3} \\ r_{\rho_4} \\ r_{c_1} \\ r_{c_2} \\ r_{c_3} \\ r_{c_4} \end{bmatrix}, \tag{12}$$

the impedance matrix becomes

$$\mathcal{A}(r) = \begin{bmatrix} \omega^2 r^1 r^5 - 2r^1 - r^2 - r^3 & r^2 & r^3 & 0 \\ r^2 & \omega^2 r^2 r^6 - 2r^2 - r^4 & 0 & r^4 \\ r^3 & 0 & \omega^2 r^3 r^7 - 2r^3 - r^4 & r^4 \\ 0 & r^4 & r^4 & \omega^2 r^4 r^8 - 2r^4 \end{bmatrix},$$

which is the new $a_{ij}(r)$ matrix from which the wavefield is computed:

$$u_i = a_{ij}^{-1}(r^\mu) f_j. \tag{13}$$

The gradient of a least-squares objective function

The frequency-domain full waveform inversion objective function is

$$\Psi(s^\mu) = \frac{1}{2} \sum_\omega \sum_j ||\mathbf{u} - \mathbf{d}||_2^2 = \sum_\omega \sum_j \Phi(s^\mu), \tag{14}$$

where $|| \cdot ||_2^2$ is the square of the 2-norm, and $\mathbf{u}$ is the simulated wavefield at frequency ω for the jth source f_i. The data vector $\mathbf{d}$ contains observations of the seismic wave field at points corresponding to the solution nodes, with the difference vector being set to zero at any node where no datum exists.* We will analyze the single source/frequency function $\Phi(s^\mu)$; all later formulas can be generalized by adding the two sums as a prefix. A full waveform inversion update is based on the gradient of Φ evaluated at a starting point s_0^μ. This is computable via

$$\frac{\partial \Phi}{\partial \mathbf{s}} = \text{Re} \left[\mathbf{u}^T \left(\frac{\partial \mathcal{A}}{\partial \mathbf{s}} \right)^T \mathcal{A}^{-1} (\mathbf{u}^* - \mathbf{d}^*) \right]. \tag{15}$$

Assuming that updates include contributions from both positive and negative frequencies, we can drop the 'Re' symbol. Doing so, and switching to indicial notation, we have

$$\frac{\partial \Phi}{\partial s^\mu} = \Phi_{,\mu}(s) = u_i \left(\frac{\partial a_{ij}(s)}{\partial s^\mu} \right) a_{jk}^{-1}(s) (u_k^* - d_k^*) \tag{16}$$

---

*More generally $\mathbf{u}$ can be replaced with $\mathbf{Ru}$ where $\mathbf{R}$ is a sampling matrix that matches the full wavefield to the measured data.

in the modulus-density parameterization, and

$$\frac{\partial \Phi}{\partial r^\mu} = \Phi_{,\mu}(r) = u_i \left(\frac{\partial a_{ij}(r)}{\partial r^\mu} \right) a_{jk}^{-1}(r) \left(u_k^* - d_k^* \right), \tag{17}$$

in the velocity-density parameterization.

Descent-based updates

First order derivative information about the objective function can be augmented with second order information via the Gauss-Newton Hessian, which in indicial notation is

$$\phi_{\mu\nu}(s) = \left(u_i \frac{\partial a_{ij}(s)}{\partial s^\mu} a_{jk}^{-1}(s) \right) \left(a_{kl}^{-1}(s) \frac{\partial a_{lm}(s)}{\partial s^\nu} u_m \right)^*, \tag{18}$$

in the modulus-density parameterization, and

$$\phi_{\mu\nu}(r) = \left(u_i \frac{\partial a_{ij}(r)}{\partial r^\mu} a_{jk}^{-1}(r) \right) \left(a_{kl}^{-1}(r) \frac{\partial a_{lm}(r)}{\partial r^\nu} u_m \right)^*, \tag{19}$$

in the velocity-density parameterization. With these quantities, two different iterative updating formulas can be set up. One is based entirely on the gradient, and is referred to as a *steepest-descent* update

$$\Delta s_{\text{sd}}^\mu = -\alpha \Phi^{,\mu}(s) \tag{20}$$

and the other includes the Gauss-Newton Hessian operator, and is referred to as a *Gauss-Newton* update:

$$\Delta s_{\text{gn}}^\mu = -\alpha \left(\phi_\lambda^\mu \right)^{-1}(s) \Phi^{,\lambda}(s), \tag{21}$$

where $\Phi^{,\mu}(s) = g^{\mu\nu} \Phi^{,\mu}(s)$ and $\phi_\lambda^\mu(s) = g^{\mu\nu} \phi_{\nu\lambda}(s)$. In both cases the scalar α is determined through a line search (e.g., Nocedal and Wright, 1999).

Vector components of the FWI gradient

Let us apply the calculations in (16) to our small prototype example. The gradients in the two parameterizations in (16) and (17) can be written as

$$\Phi_{,\mu}(s) = u_i(s) \left(\frac{\partial a_{ij}(s)}{\partial s^\mu} \right) v_j(s), \quad \Phi_{,\mu}(r) = u_i(r) \left(\frac{\partial a_{ij}(r)}{\partial r^\mu} \right) v_j(r), \tag{22}$$

where $v_j(s) = a_{jk}^{-1}(s) \left(u_k^* - d_k^* \right)$ is the back-propagated residual wavefield (e.g., Pratt, 1999), and where we have emphasized the dependence of the u_i and v_j on the s^μ or r^μ at which the derivatives are being evaluated. In Appendix A, the derivatives of the elements of $\mathcal{A}(s)$ and $\mathcal{A}(r)$ with respect to each element of s^μ and r^μ are given as a set of 4×4 matrices, one for each of $8\ \mu$ values. Extracting the i,jth component of each and arranging

these in a column vector gives the $\partial a_{ij}/\partial s^\mu$. For instance, in the s system at the first and third diagonals ($i = j = 1$ and $i = j = 3$) we have

$$\frac{\partial a_{11}}{\partial s^\mu} = \begin{bmatrix} -2 \\ -1 \\ -1 \\ 0 \\ \omega^2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \frac{\partial a_{33}}{\partial s^\mu} = \begin{bmatrix} 0 \\ 0 \\ -2 \\ -1 \\ 0 \\ 0 \\ \omega^2 \\ 0 \end{bmatrix}, \tag{23}$$

whereas in the r system we have

$$\frac{\partial a_{11}}{\partial r^\mu} = \begin{bmatrix} \omega^2 r^5 - 2 \\ -1 \\ -1 \\ 0 \\ \omega^2 r^1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \frac{\partial a_{33}}{\partial r^\mu} = \begin{bmatrix} 0 \\ 0 \\ \omega^2 r^7 - 2 \\ -1 \\ 0 \\ 0 \\ \omega^2 r^3 \\ 0 \end{bmatrix}. \tag{24}$$

These vectors are then linearly combined with the weights u_i and v_j to create the full gradient vector.

RE-PARAMETERIZATION OF FWI AS A COORDINATE TRANSFORM

Re-parameterization between medium property types

We will now show that the change of parameters from modulus-density to velocity-density (which we will take as a proxy for general changes from one set of parameter classes to another) is formally a transformation between rectilinear but non-Cartesian coordinate systems. From (10) we have for small displacements in density and modulus

$$\delta s_\rho = \delta r_\rho, \text{ and } \delta s_\kappa = r_c \delta r_\rho + r_\rho \delta r_c. \tag{25}$$

Applying these to terms in the prototype 2×2 system, using the numbering system in (7), we obtain a rule for how displacements in r go over into displacements in s:

$$\begin{bmatrix} \delta s^1 \\ \delta s^2 \\ \delta s^3 \\ \delta s^4 \\ \delta s^5 \\ \delta s^6 \\ \delta s^7 \\ \delta s^8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ r^5 & 0 & 0 & 0 & r^1 & 0 & 0 & 0 \\ 0 & r^6 & 0 & 0 & 0 & r^2 & 0 & 0 \\ 0 & 0 & r^7 & 0 & 0 & 0 & r^3 & 0 \\ 0 & 0 & 0 & r^8 & 0 & 0 & 0 & r^4 \end{bmatrix} \begin{bmatrix} \delta r^1 \\ \delta r^2 \\ \delta r^3 \\ \delta r^4 \\ \delta r^5 \\ \delta r^6 \\ \delta r^7 \\ \delta r^8 \end{bmatrix}. \tag{26}$$

This defines a transformation from the r system to the s system. In our notation we have implied (by using superscript indices) that these displacements are contravariant components of vectors. Assuming this the matrix above defines the rule for transformation of contravariant vector components from the s system to the r system:

$$\delta r^\mu = t^\mu_\nu \delta s^\nu. \tag{27}$$

If this holds, it must then be true that covariant vectors transform from s to r with the transpose of the operator for transforming contravariant vectors from r to s. If the two re-parameterized FWI gradients are so related, the issue will be settled. Examining the gradient in the s system,

$$\Phi_{,\mu}(s) = u_i(s) \left(\frac{\partial a_{ij}(s)}{\partial s^\mu} \right) v_j(s), \tag{28}$$

we observe that all three terms depend on the particular s^μ at which the gradient is being formed. However, $u_i(s)$ and $v_j(s)$ at any i,j pair are scalar functions of s^μ, and so $u_i(s) = u_i(r)$ and $v_j(s) = v_j(r)$ under the transform. So, the question of the transformation of $\Phi_{,\mu}$ rests on the central term of the three. There are 10 independent vectors $\partial a_{ij}(s)/\partial s^\mu$ in our 2×2 problem, but let us exemplify the transformation with the first and third diagonals already set out in (2). The first diagonal produces, when operated on by the transpose of the matrix in (26),

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & r^5 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & r^6 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & r^7 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & r^8 \\
0 & 0 & 0 & 0 & r^1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & r^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & r^3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & r^4
\end{bmatrix}
\begin{bmatrix}
-2 \\ -1 \\ -1 \\ 0 \\ \omega^2 \\ 0 \\ 0 \\ 0
\end{bmatrix}
=
\begin{bmatrix}
\omega^2 r^5 - 2 \\ -1 \\ -1 \\ 0 \\ \omega^2 r^1 \\ 0 \\ 0 \\ 0
\end{bmatrix},
\tag{29}
$$

and the third produces

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & r^5 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & r^6 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & r^7 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & r^8 \\
0 & 0 & 0 & 0 & r^1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & r^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & r^3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & r^4
\end{bmatrix}
\begin{bmatrix}
-2 \\ -1 \\ -1 \\ 0 \\ \omega^2 \\ 0 \\ 0 \\ 0
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ \omega^2 r^7 - 2 \\ -1 \\ 0 \\ 0 \\ \omega^2 r^3 \\ 0
\end{bmatrix}.
\tag{30}
$$

Comparing these with (3), which were obtained from direct differentiation, we confirm that the transformation rule operates correctly. From this we conclude (by a significant, but warranted, extension from the current example) that model space re-parameterizations between medium property types are formally transforms between oblique-rectilinear coordinate systems overlain on model space, i.e., that

$$\delta r^\mu = t^\mu_\nu \delta s^\nu \quad \text{implies} \quad \Phi_{,\nu}(s) = t^\mu_\nu \Phi_{,\mu}(r), \tag{31}$$

and vice versa.

Re-parameterization for model-space preconditioning

We will now show that the changes of variable involved in model-space preconditioning (e.g., Fomel and Claerbout, 2003; Guitton et al., 2012) are also coordinate transforms between non-Cartesian, rectilinear coordinate systems. The idea is based on a note by Harlan (1994), who was discussing the refinement of an approach to tomographic inversion developed by Bube and Langan (1994). We assume an objective function of the form

$$\Phi(\mathbf{m}) = \frac{1}{2}||\mathbf{u}(\mathbf{m}) - \mathbf{d}||_2^2 + \alpha\frac{1}{2}||\mathbf{T}^{-1}\mathbf{m}||_2^2, \tag{32}$$

with $\mathbf{T}^{-1}$ being a regularization term which "roughens" the model, and updates that are parallel to the gradient

$$\frac{\partial\Phi}{\partial\mathbf{m}} = \frac{\partial\mathbf{u}^T}{\partial\mathbf{m}}\left(\mathbf{u}(\mathbf{m}) - \mathbf{d}\right)^* + \alpha\left(\mathbf{T}^{-1}\right)^T\mathbf{Tm}, \tag{33}$$

with $\cdot^T$ meaning transpose or conjugate transpose as needed. Bube and Langan recommended that α be varied with iteration, taking on large values at the beginning of the procedure and shrinking as iterations proceeded. Harlan pointed out that this process of continuation achieves its goal more effectively under the change of variables $\mathbf{m} = \mathbf{Tm}'$, where $\mathbf{T}^{-1}\mathbf{T} \approx \mathbf{I}$.

Expressing these basic equations in covariant notation, and letting the un-primed model vector $\mathbf{m}$ be associated with s^μ, we have for the objective function

$$\Phi(s) = \frac{1}{2}\left(u_i(s) - d_i\right)\left(u_i^*(s) - d_i^*\right) + \alpha(t^{-1})_\lambda^\gamma s^\lambda g_{\gamma\nu}(t^{-1})_\sigma^\nu s^\sigma, \tag{34}$$

and the gradient

$$\Phi_{,\mu}(s) = \frac{\partial u_i}{\partial s^\mu}\left(u_i^*(s) - d_i^*\right) + \alpha(t^{-1})_\mu^\gamma g_{\gamma\nu}(t^{-1})_\sigma^\nu s^\sigma. \tag{35}$$

If the change of variables as recommended by Harlan is the contravariant coordinate transform $s^\mu = t_\nu^\mu r^\nu$, where t_ν^μ is approximately the inverse of $(t^{-1})_\mu^\nu$, it follows that the components of the gradient transform using the covariant rule:

$$\Phi_{,\mu}(r) = t_\mu^\nu\Phi_{,\nu}(s) \approx \frac{\partial u_i}{\partial r^\mu}\left(u_i^*(s) - d_i^*\right) + \alpha g_{\mu\nu}r^\nu, \tag{36}$$

with the $\approx$ sign reflecting the fact that the two operators not be exact inverses of one another, i.e., $t_\lambda^\mu(t^{-1})_\nu^\lambda \approx \delta_\nu^\mu$. Notice that if we leave the Jacobian in terms of the s system, i.e.,

$$\Phi_{,\mu}(r) = t_\mu^\nu\Phi_{,\nu}(s) = t_\mu^\nu\frac{\partial u_i}{\partial s^\mu}\left(u_i^*(s) - d_i^*\right) + \alpha g_{\mu\nu}r^\nu, \tag{37}$$

when we transform back to vector notation

$$\frac{\partial\Phi}{\partial\mathbf{m}'} = \mathbf{T}^T\frac{\partial\mathbf{u}^T}{\partial\mathbf{m}}\left(\mathbf{u}\left(\mathbf{Tm}'\right) - \mathbf{d}\right)^* + \alpha\mathbf{m}', \tag{38}$$

we reproduce the key result in the note (Harlan's equations 7-8). Thus we conclude that model space preconditioning is also in essence the result of carrying out a covariant coordinate transformation between rectilinear but oblique axes.

DESIGN OF FWI COORDINATE SYSTEMS BASED ON THE HESSIAN

With the re-parameterization of FWI thus characterized in terms of transformation between rectilinear coordinate systems in place, we next examine the possibility of designing alternative transforms with favourable features. In particular, as set out in a companion report (Innanen, 2020a), we consider systems for which the Hessian matrix is a Kronecker delta, within which there is a low probability of parameter cross-talk, and steepest-descent update directions are parallel to the normally much more efficient Newton update directions.

NUMERICS

An exhaustive numerical analysis of the re-parameterization proposed in the previous sections, and elsewhere in this report, is underway, but that will require the examination of many approaches to pre-selecting transform matrix entries, the analysis of various types of multiparameter FWI, regularizations, etc. Here we will restrict ourselves to one basic application in order to demonstrate that the idea works in essence, and that its application produces distinct improvements on steepest-descent optimization.

A small FWI problem optimized with steepest descent

We adopt a small scalar acoustic example, consisting of a background velocity model which increases linearly with depth, interrupted by a high velocity anomaly in the centre. The background linear velocity model will be assumed to be known, and will be used as the initial model. The geological model is illustrated in image form in Figure 2a, with a central profile extracted and plotted in Figure 2c. The initial model is likewise plotted in image form in Figure 2b, with a central profile plotted in Figure 2d.
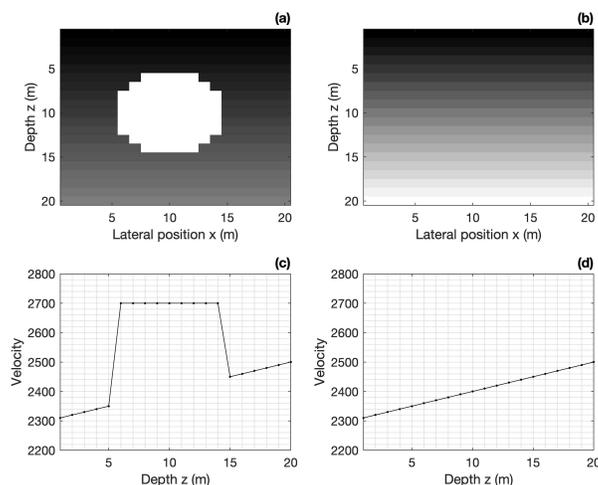
FIG. 2. A small (20x20 cell) model is adopted as a backdrop for numerical testing. (a) The geology is a linearly increasing P-wave velocity interrupted by a high velocity zone; (b) the background linear velocity is adopted as the initial model. In (c)-(d) profiles are extracted from the centre to enable more precise monitoring of the FWI convergence.

We treat this problem first with a "reference" FWI procedure, comprising a steepest descent optimization of a frequency-domain misfit function without regularization. In Figure

3a source (circle) and receiver (dot) positions are illustrated over the geological model to be recovered. A multiscale inversion is set up, involving 20 iterations, each using a different set of 5 frequencies, with the frequencies fanning out over a wider range at each iteration (Keating and Innanen, 2019). At iteration 1, a single frequency of 1Hz is used, and at iteration 20, the frequencies span the 1-35Hz band. The exact frequencies used are enumerated in Appendix B.
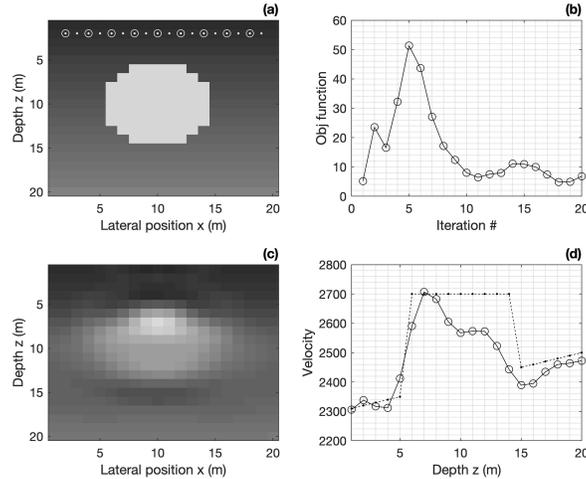
FIG. 3. Benchmark steepest-descent optimization example. (a) The true model with source (circle) and receiver (dot) locations; (b) data misfit per iteration; (c) final results after 20 iterations; (d) profile of actual result versus steepest descent result.

Data misfit per iteration is plotted in Figure 3b for reference; because new frequencies are added with each iteration, the plot is generally non-decreasing especially in the early iterations. In Figures 3c-d the final results of the inversion are plotted, in image and profile form respectively. The structure and velocity of the anomaly are beginning to appear, though bounded by some incorrectly low velocity regions, and with an underestimated velocity in the deeper portions. The limited acquisition geometry, even supported by 1Hz data, tends to produce dimmer velocities to the left and right of the centre of the anomaly as well. We have therefore in hand a good benchmark example, which works reasonably well but also exhibits challenges that second order Hessian information is known to at least partly address.

Coordinate transforms

Our benchmark problem is small enough to allow us to explicitly calculate the Hessian (which we emphasize is not necessary in order to use the transform approach). For an example, at the 10th iteration of the 20 iteration sequence in the benchmark optimization above, we pause and compute the Hessian matrix; it is plotted in Figure 4. For a 20x20 cell example, this matrix is 400x400 in size, and made up of 20x20 blocks of size 20x20 each.

The matrix is poorly conditioned (in this case with a condition number of roughly 10^6). To make any further use of it we add a stabilizer:

$$\mathbf{H}_{\text{stabilized}} = \mathbf{H} + \lambda\mathbf{I}, \tag{39}$$
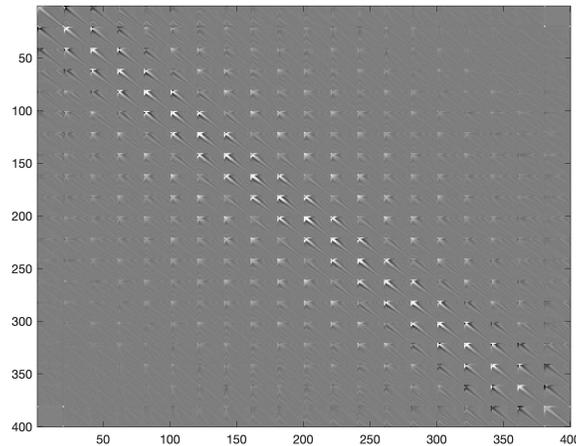
FIG. 4. The Hessian matrix calculated during the 10th iteration of the benchmark steepest descent optimization illustrated in Figure 3.

with λ chosen to be roughly 1% of the norm of the matrix.

At this point in the problem we could invert this stabilized Hessian matrix, and carry out a Gauss-Newton optimization, which we would expect to address at least some of the illumination issues encountered in Figure 3. Instead, we will undertake to design a transform operator t_ν^μ following the approach set out in a companion report (Innanen, 2020b). This requires a pre-selection of the lower triangular elements of a 400x400 matrix, followed by the determination of the remaining diagonal and upper triangular elements. For simplicity, we select the lower elements to be zeros. After this selection has taken place there now exists only one transformation matrix for which the transformed Hessian is the Kronecker delta.

The matrix emerging from the design process is plotted in Figure 5. The nil lower triangular aspect is obvious to the eye, along with a large diagonal value and banding in the upper triangular region.

Standard and transformed optimization

In order to facilitate a controlled analysis of the use of the transformation domain for optimization, we will repeat the 20 iteration steepest-descent inversion carried out above, but interrupt it at a particular iteration, optimize in the transform domain for that one iteration, and then return to our standard steepest descent approach. From Figure 3b it is evident that at and around the 5th iteration are the largest changes in data misfit, so we select iteration 5 for this experiment.

In Figure 6 the results of the inversion starting from the initial point and ending at the 4th iteration are plotted. The general low wavenumber positive trend caused by the anomaly is present, but little detail has been incorporated as yet. From this point our experiment diverges, and two independent results are produced. The first is a repeat of the original numerical experiment; the second involves carrying out the optimization at the
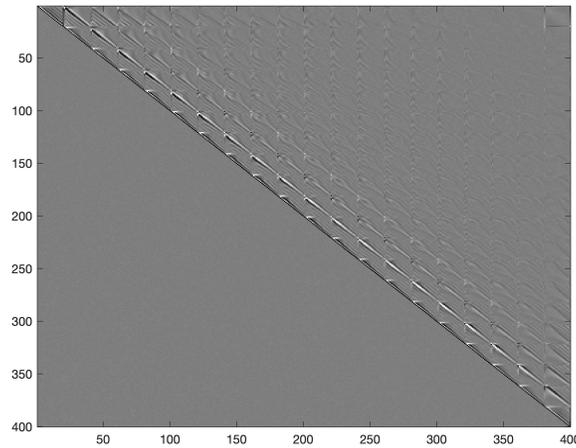
FIG. 5. The unique transformation matrix with zero entries below the diagonal and under which the Hessian operator at the 10th iteration above is a Kronecker delta.
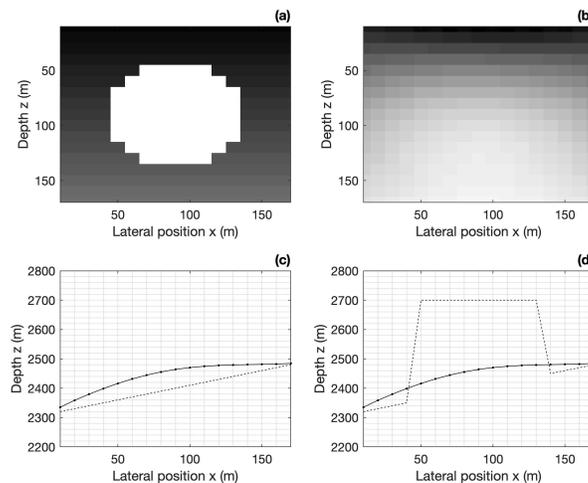
FIG. 6. Steepest descent optimization is carried out to iteration 4 and paused. (a) Actual simulated geological model; (b) updated model in image form at 4th iteration; (c) central profile of updated model plotted against initial model; (d) central profile of updated model plotted against actual model.

5th iteration in the coordinate system defined by the transformation operator illustrated in Figure 5.

In Figure 7 the two columns (a-c) and (b-d) illustrate the two results at the end of the 5th iteration. In panels (a)-(c) is the original steepest-descent optimization, in image form and profile form respectively; in (b)-(d) the result optimized in the transformed space is plotted. The procedure for the transformed optimization is simply to move to the new space, carry out a steepest descent update in that space, and then move back. Visually, it appears that the transformed optimization has had an effect on the amplitude of the update, evidently reflecting a boosting of amplitudes to compensate for illumination, a feature known to be present in the Hessian operator.

The result of the single transform-optimized iteration propagates through to the final result. In Figures 8a-c and b-d, as before, the standard and transform-optimized results after
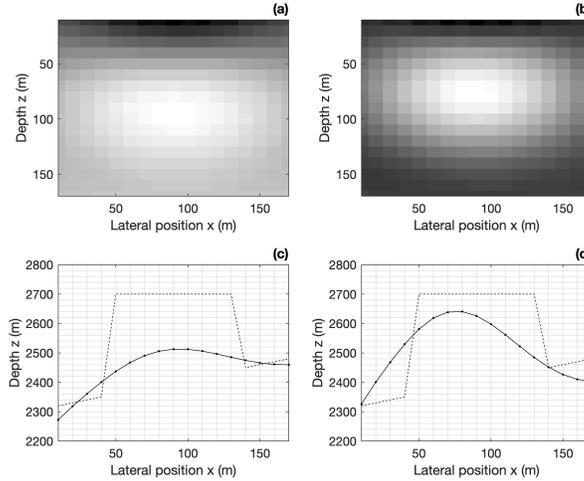
FIG. 7. A comparison of optimizations carried out at the 5th iteration, using (a)-(c) a steepest-descent update in standard coordinates; and (b)-(d) a steepest descent update in the transformed coordinate system (and then returning).
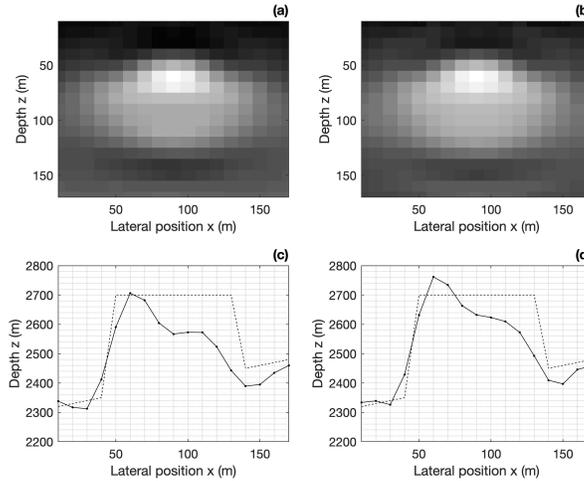
FIG. 8. (a)-(c) 20 iterations, after a standard 5th iteration; (b)-(d) 20 iterations after a transform optimization at the 5th iteration. Top panels: full images; bottom panels: profiles.

20 iterations are plotted in image and profile form. The boost or illumination compensation remains visible.

CONCLUSIONS

The re-parameterizations we encounter regularly in seismic full waveform inversion can all be formally identified as coordinate transforms between suitably general coordinate systems. For instance, changes-of-variable between elastic/acoustic parameter classes, and the variable changes underlying the model space pre-conditioning methods of Harlan, Claerbout and Fomel, can both be shown to be transforms between oblique-rectilinear coordinate systems when covariant notation is employed. With transformation rules in place for updates and displacements (contravariant vectors), and sensitivities and gradients (covariant vectors), as well as objective functions (scalars) and Hessian operators (2nd rank covariant tensors), we may then seek new and potentially useful re-parameterizations. Coordinate

FIG. 9. (a-d) Various vertical profiles through the final results, through basic steepest descent (black) and through transformed steepest descent (blue).

systems within which the Hessian is close to the identity have vastly improved optimization properties. Using a newly-derived numerical procedure for designing transformation matrices mapping to such systems, we provide an early examination of re-parameterized FWI.

APPENDIX A

In the modulus-density parameterization, the 8 derivatives of the impedance matrix are

$$\frac{\partial a_{ij}(s)}{\partial s^1} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(s)}{\partial s^2} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(s)}{\partial s^3} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\frac{\partial a_{ij}(s)}{\partial s^4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & 1 & -2 \end{bmatrix}, \quad \frac{\partial a_{ij}(s)}{\partial s^5} = \begin{bmatrix} \omega^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(s)}{\partial s^6} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \omega^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\frac{\partial a_{ij}(s)}{\partial s^7} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \omega^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(s)}{\partial s^8} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega^2 \end{bmatrix}.$$

In the velocity-density parameterization, the derivatives are

$$\frac{\partial a_{ij}(r)}{\partial r^1} = \begin{bmatrix} \omega^2 r^5 - 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(r)}{\partial r^2} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & \omega^2 r^6 - 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(r)}{\partial r^3} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & \omega^2 r^7 - 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\frac{\partial a_{ij}(r)}{\partial r^4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & 1 & \omega^2 r^8 - 2 \end{bmatrix}, \quad \frac{\partial a_{ij}(r)}{\partial r^5} = \begin{bmatrix} \omega^2 r^1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(r)}{\partial r^6} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \omega^2 r^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\frac{\partial a_{ij}(r)}{\partial r^7} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \omega^2 r^3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \frac{\partial a_{ij}(r)}{\partial r^8} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega^2 r^4 \end{bmatrix}.$$

APPENDIX B

The basic multiscale procedure involved the simultaneous use of 5 frequencies per iteration, with the frequencies being spread over an increasing range as iterations proceeded. The low end and high end f values are specified; at iteration 1 the f's are all equal to the low value and at the maximum iteration they are evenly spread between the low and high value; the high end f value is arrived at through even steps as the iterations proceed. The specific values this implies are included below.

Iteration	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	f_4 (Hz)	f_5 (Hz)
1	1.0	1.0	1.0	1.0	1.0
2	1.0	1.4	1.9	2.3	2.8
3	1.0	1.9	2.8	3.7	4.6
4	1.0	2.3	3.7	5.0	6.4
5	1.0	2.7	4.6	6.4	8.2
6	1.0	3.2	5.5	7.7	9.9
7	1.0	3.7	6.4	9.0	11.7
8	1.0	4.1	7.3	10.4	13.5
9	1.0	4.6	8.2	11.7	15.3
10	1.0	5.0	9.0	13.0	17.1
11	1.0	5.5	9.9	14.4	18.9
12	1.0	5.9	10.8	15.8	20.7
13	1.0	6.4	11.7	17.1	22.5
14	1.0	6.8	12.6	18.4	24.3
15	1.0	7.3	13.5	19.8	26.1
16	1.0	7.7	14.4	21.1	27.8
17	1.0	8.2	15.3	22.4	29.6
18	1.0	8.6	16.2	23.8	31.4
19	1.0	9.1	17.1	25.2	33.2
20	1.0	9.5	18.0	26.5	35.0

Table 1. Frequency groups for multiscale FWI updating.

REFERENCES

Anagaw, A., and Sacchi, M. D., 2018, Model parametrization strategies for Newton-based acoustic full waveform inversion: Journal of Applied Geophysics, **157**, 23–36.

Bube, K., and Langan, R. T., 1994, A continuation approach to regularization for traveltime tomography: SEG Expanded Abstracts.

Fomel, S., and Claerbout, J. F., 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: Geophysics, **68**, 577–588.

Guitton, A., Ayeni, G., and Diaz, E., 2012, Constrained full-waveform inversion by model reparameterization: Geophysics, **77**, No. 2, R117–R127.

Harlan, W. S., 1994, Regularization by model reparameterization: http:// billharlan. com/ pub/ papers/ regularization/ regularization.html.

Hu, Q., and Innanen, K. A., 2020, Elastic FWI with rock physics constraints: CREWES Annual Report, **32**.

Innanen, K. A., 2020a, Rep 2. Model re-parameterization via misfit-based coordinate transforms: CREWES Annual Report, **32**.

Innanen, K. A., 2020b, Rep 5. Numerical procedures for computing constrained coordinate transforms: CREWES Annual Report, **32**.

Keating, S. D., and Innanen, K. A., 2019, Parameter crosstalk and modeling errors in viscoacoustic seismic full-waveform inversion: Geophysics, **84**.

Keating, S. D., and Innanen, K. A., 2020, Simultaneous recovery of source locations, moment tensors and subsurface models in 2D FWI: CREWES Annual Report, **32**.

Nocedal, J., and Wright, S. J., 1999, Numerical Optimization, Springer Series in Operations Research: Springer-Verlag.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model: Geophysics, **64**, No. 3, 888–901.