Multiparameter full-waveform inversion with a SBP-SAT discretization of the first order acoustic wave equation

Joseph Jennings, Martin Almquist and Eric Dunham

ABSTRACT

We develop the theory for implementing multiparameter full-waveform inversion using high-order accurate summation-by-parts finite difference operators and weak enforcement of boundary conditions. With these discrete operators, we derive the semi-discrete adjoint equations and gradients that closely mimic those of the continuous problem. We provide a numerical example in which we estimate the source time function from synthetic pressure data. We anticipate that this formulation will be useful for complicated modeling scenarios such as modeling point sources directly on the ocean bottom interface.

INTRODUCTION

As imaging targets become increasingly complex and more information is desired from the subsurface, more advanced acquisition, imaging and modeling algorithms are required. One such example of this is the increase in the desire to estimate the elastic parameters of the subsurface. For marine data this requires placing either ocean-bottom nodes or cables on the seafloor to record the elastic wavefield. Then when Zoeppritz approximation fails, methods such as elastic full-waveform inversion (FWI) are required to obtain reliable estimates of subsurface elastic parameters (Biondi et al., 2016). To perform elastic FWI, in addition to accurate starting models, accurate finite-difference modeling schemes are needed to ensure adequate data-fitting.

FWI is a partial differential equation (PDE)-constrained optimization problem that is commonly minimized via gradient-based methods (Bradley, 2013). The gradient calculation of the FWI objective function requires a solution to what is known as the forward (PDE) and an adjoint PDE. Using a standard least-squares misfit, the source of the adjoint PDE is the mismatch between the predicted data and the actual data and is injected at the receiver locations. For the ocean-bottom acquisition setup, this requires solving the adjoint PDE with a source directly on a fluid-solid interface. Handling this interface in the correct manner has not been well-studied in the context of FWI. Ober et al. (2016) use a discontinuous-galerkin scheme and impose that the normal component of velocity is continuous across the boundary and that the tangential component of the traction along the boundary is zero. While there have

been several case studies in which elastic FWI has been performed with ocean-bottom multicomponent data,((Sears et al., 2010; Prieux et al., 2013; Ober et al., 2016; Alves, 2017)) none provide a rigorous treatment of how to solve the elastic wave equation with a source at the fluid-solid interface and a finite-difference discretization.

The work herein describes how to systematically perform complex forward and adjoint modeling with a finite difference scheme known as summation-by-parts (SBP) with the simultaneous approximation term method that weakly enforces the boundary conditions. It has been successfully used in modeling situations with complex geometry that require accurate solutions to the PDE at the boundary (Lotto and Dunham, 2015; Duru and Dunham, 2016; Karlstrom and Dunham, 2016). Additionally, the scheme is energy stable and can be generalized beyond just cartesian grids allowing for an undulating ocean bottom. While this SBP finite difference scheme is not new and has been used in solving a FWI for estimating a moment tensor source (Sjögreen and Petersson, 2014), as well as reverse-time migration (Wang et al., 2017), what is new in this work is that we derive the adjoint equations and gradients for SBP operators that operate on staggered grids. We show that the use of these operators lead to self-adjoint spatial derivative operators. Moreover, we demonstrate that in using this scheme, we achieve dual-consistency which indicates that the discretization of the continuous adjoint equations is consistent with the discrete derivation of the adjoint equations (Berg and Nordström, 2012).

We begin with first setting up the FWI optimization problem for the continuous case. We derive the adjoint equations and the gradient for the medium parameters and source for a first-order acoustic wave equation. We then discretize the forward equations in space using a SBP discretization and derive the adjoint equations and gradient for the semi-discrete problem in the case in which we do not consider the boundaries. We show that as for the continuous problem, the semi-discrete differential operators are self-adjoint. We also show this for the scenario in which we have a free-surface boundary condition. We then show a 1D numerical example in which we estimate a source time function from synthetic pressure data.

ADJOINT EQUATIONS AND GRADIENT FOR THE CONTINUOUS PROBLEM

We first derive the adjoint equations and gradient for the continuous problem. We do this by first introducing the forward PDE which is a system of first-order equations that describe wave propagation in an acoustic medium. While we limit our derivations to the case of one-dimension, these derivations can be generalized to 3-dimensions.

The governing equations for our problem are the following momentum and mass

balance

$$\rho \frac{\partial v}{\partial t} + \frac{\partial p}{\partial x} = 0 \tag{1}$$

$$\frac{1}{K}\frac{\partial p}{\partial t} + \frac{\partial v}{\partial x} = f(t)\delta(x - x_s), \tag{2}$$

with the following initial condition

$$v(x,0) = 0, \quad p(x,0) = 0, \tag{3}$$

and where we have not specified any specific boundary condition. These equations are the governing equations for waves propagating in a 1D acoustic medium with a wave-speed $c = \sqrt{K/\rho}$. v denotes particle velocity, p is the pressure, ρ is density, K is bulk modulus and $f(t)$ is a source time function positioned at the source location x_s. For this example we set $x \geq 0$ and enforce boundary conditions on $x = 0$. We will consider the forward problem for the free-surface boundary condition ($p(0,t) = 0$), a rigid boundary condition ($v(0,t) = 0$) and non-reflecting boundary conditions ($p(0,t) + \rho c v(0,t) = 0$).

Now we desire to look at the gradient and adjoint equations for the FWI problem in which we desire to estimate the medium parameters, namely ρ and K, and the source time function $f(t)$. The PDE-constrained optimization problem can be written as

$$\begin{aligned}
&\underset{\rho(x),K(x),f(t)}{\text{minimize}} \quad F(v,p,\rho,K,f(t)), \quad \text{where} \\
&F(v,p,\rho,K,f(t)) = \frac{w_1}{2}\int_0^T \left(v(x_r,t) - v_{\text{data}}(t)\right)^2 dxdt \\
&\qquad\qquad + \frac{w_2}{2}\int_0^T \left(p(x_r,t) - p_{\text{data}}(t)\right)^2 dxdt \qquad (4) \\
&\text{subject to} \quad \rho\dot{v} + p_x = 0, \\
&\qquad\qquad \frac{1}{K}\dot{p} + v_x - f(t)\delta(x - x_s) = 0,
\end{aligned}$$

where Ω indicates the entire spatial domain. Note that we have introduced the additional constant weights w_1 and w_2 which control the contribution of residuals from pressure and particle velocity respectively. In the case in which we have only p_{data} or only v_{data}, then either w_1 or w_2 is zero. In the case in which we have both, then we might choose $w_1 = \rho_r$ and $w_2 = K_r^{-1}$ where the subscript r indicates the material property at the receiver location. This choice of weights would result in minimizing the energy of the residuals. Another possible choice could be $w_1 = 1$ and $w_2 = z_r$ where z_r is the acoustic impedance at the receiver location. Also note that while this functional is stated for a single receiver, it can be generalized in a straightforward manner for an arbitrary number of receivers by adding additional terms to the misfit

functional. We now introduce Lagrange multipliers $\lambda_1(x,t)$ and $\lambda_2(x,t)$ which turn this constrained optimization problem into an unconstrained problem

$$\begin{aligned}\mathcal{L} = &\frac{w_1}{2}\int_0^T \left(v(x_r,t) - v_{\text{data}}(t)\right)^2 dxdt \\ &+ \frac{w_2}{2}\int_0^T (p(x_r,t) - p_{\text{data}}(t))^2 dxdt + \int_0^T\int_\Omega \lambda_1\left(\rho\dot{v} + p_x\right) dxdt \\ &+ \int_0^T\int_\Omega \lambda_2\left(\frac{1}{K}\dot{p} + v_x - f(t)\delta(x-x_s)\right) dxdt. \end{aligned} \tag{5}$$

We first desire to calculate first variation of the functional denoted as $\delta\mathcal{L}$. The first variation is the change in the functional $\mathcal{L}$ when the parameters of interest are changed ($\delta\rho$, δK and $\delta f(t)$). For our problem, we can write this as

$$\delta\mathcal{L} = \int_\Omega \frac{\delta\mathcal{L}}{\delta\rho(x)}\delta\rho(x)dx + \int_\Omega \frac{\delta\mathcal{L}}{\delta K(x)}\delta K(x)dx + \int_0^T \frac{\delta L}{\delta f(t)}\delta f(t)dt \tag{6}$$

where the terms $\delta\mathcal{L}/\delta\rho(x)$, $\delta\mathcal{L}/\delta K(x)$ and $\delta\mathcal{L}/\delta f(t)$ are known as functional derivatives. When using the adjoint method in the geophysics community, they are also referred to as sensitivity kernels (Liu and Tromp, 2008). Due to the fact that the wavefields v and p are dependent on these material properties and source time function, we will also have terms δp and δv in our calculation of the first variation. The first variation of equation 5 can be expressed as

$$\begin{aligned}\delta\mathcal{L} = &w_1\int_0^T\int_\Omega \left(v(x,t) - v_{\text{data}}(x_r,t)\right)\delta(x-x_r)\delta v dxdt \\ &+ w_2\int_0^T\int_\Omega (p(x,t) - p_{\text{data}}(x_r,t))\delta(x-x_r)\delta p dxdt \\ &+ \int_0^T\int_\Omega \lambda_1\left(\delta\rho\dot{v} + \delta\dot{v}\rho + \delta p_x\right) dxdt \\ &+ \int_0^T\int_\Omega \lambda_2\left(-K^{-2}\delta K\dot{p} + K^{-1}\delta\dot{p} + \delta v_x - \delta f(t)\delta(x-xs)\right) dxdt. \end{aligned} \tag{7}$$

Now, in order to avoid computing the variations in the wavefields δv and δp we first integrate by parts in space and in time all terms associated with spatial and temporal derivatives of δv and δp. This will switch the derivatives on δp and δv to λ_1 and

λ_2 and also introduce boundary terms. Performing this integration by parts and grouping terms associated with δp and δv we obtain the following expression for the augmented functional

$$\begin{aligned}
\delta\mathcal{L} = &\int_0^T \int_\Omega \left(w_1 \left(v - v_{\text{data}}\right) \delta(x - x_r) - \rho\dot{\lambda_1} - \lambda_{2x} \right) \delta v \\
&+ \int_0^T \int_\Omega \left(w_2 (p - p_{\text{data}}) \delta(x - x_r) - K^{-1}\dot{\lambda_2} - \lambda_{1x} \right) \delta p dx dt \\
&+ \int_\Omega \rho \left(\lambda_1(x,T)\delta v(x,T) - \lambda_1(x,0)\delta v(x,0) \right) \\
&+ \int_\Omega K^{-1} \left(\lambda_2(x,T)\delta p(x,T) - \lambda_2(x,0)\delta p(x,0) \right) dx \\
&+ \int_0^T \lambda_1(0,t)\delta p(0,t) + \lambda_2(0,t)\delta v(0,t) dt \\
&+ \int_0^T \int_\Omega \lambda_1 \delta\rho\dot{v} - K^{-2}\delta K \lambda_2 \dot{p} - \lambda_2 \delta f(t) \delta(x - x_s) dx dt
\end{aligned} \tag{8}$$

Now that we have factored out the wavefield variation terms we can impose conditions on the Lagrange multipliers in order to avoid the computation of these terms. The conditions we impose are written as follows

$$\rho\dot{\lambda}_1 + \lambda_{2x} = \; w_1(v - v_{\text{data}})\delta(x - x_r), \tag{9}$$

$$\frac{1}{K}\dot{\lambda}_2 + \lambda_{1x} = \; w_2(p - p_{\text{data}})\delta(x - x_r), \tag{10}$$

$$\lambda_1(x,T) = 0, \quad \lambda_2(x,T) = 0. \tag{11}$$

We observe that equations 9 and 10 are PDEs of the same form as equations 1 and 2 and make what are known as adjoint PDEs with terminal conditions 11. We next consider the conditions that we need to place on the Lagrange multipliers to derive the adjoint boundary conditions.

Adjoint boundary terms

We now examine the boundary terms that appear in the expression

$$\int_0^T \lambda_1(0,t)\delta p(0,t) + \lambda_2(0,t)\delta v(0,t) dt. \tag{12}$$

In order to obtain the adjoint boundary conditions from this expression, we first impose a boundary condition on $p(x,t)$ and $v(x,t)$ which then implies a boundary condition on $\delta v(x,t)$ and $\delta p(x,t)$ (Liu and Tromp, 2006) and then we derive the resulting adjoint boundary condition by setting the Lagrange multipliers to force the integrand to zero. We show this for the rigid, free surface and absorbing boundary conditions.

Rigid boundary

For the rigid boundary, we impose that $v(0,t) = 0 \Rightarrow \delta v(0,t) = 0$. Using this fact we find the adjoint boundary condition as $\lambda_1(0,t) = 0$.

Free surface

Being that the free surface boundary condition ($p(0,t) \Rightarrow \delta p(0,t) = 0$) is the complement of the rigid boundary condition, we find then that $\lambda_2(0,t) = 0$.

Absorbing boundary

For the absorbing boundary, we have that $p(0,t) + \rho c v(0,t) = 0 \Rightarrow \delta p(0,t) + \rho c \delta v(0,t) = 0$. Using this expression along with the integrand in equation 12, we find the adjoint absorbing boundary to be

$$\begin{aligned} -\lambda_1(0,t)\rho c \delta v(0,t) + \lambda_2(0,t)\delta v(0,t) &= 0 \\ \Rightarrow -\rho c \lambda_1(0,t) + \lambda_2(0,t) &= 0. \end{aligned}$$

Gradients of the continuous problem

As equations 9 and 10 form a system of PDEs with terminal conditions 11, we perform a change of temporal coordinates $\tau = T - t \Rightarrow d\tau = -dt$. This results in the following time-reversed PDEs

$$-\rho\lambda_1' + \lambda_{2x} = \ w_1(v - v_{\text{data}})\delta(x - x_r), \tag{13}$$

$$-\frac{1}{K}\lambda_2' + \lambda_{1x} = \ w_2(p - p_{\text{data}})\delta(x - x_r), \tag{14}$$

$$\lambda_1(x, \tau = 0) = 0, \quad \lambda_2(x, \tau = 0) = 0, \tag{15}$$

where the λ_1' indicates a derivative of λ_1 in τ. Additionally, we let $-\lambda_1 = \lambda_3$ which results in the following system of PDEs

$$\rho\lambda_3' + \lambda_{2x} = w_1(v - v_{\text{data}})\delta(x - x_r), \tag{16}$$

$$\frac{1}{K}\lambda_2' + \lambda_{3x} = \ - w_2(p - p_{\text{data}})\delta(x - x_r), \tag{17}$$

$$\lambda_3(x, \tau = 0) = 0, \quad \lambda_2(x, \tau = 0) = 0. \tag{18}$$

Note that this system is of the exact same form as the forward PDE shown in equations 1 and 2. Upon satisfying conditions 16-18 (solving the adjoint equations) the calculation of the first variation can be written as

$$\delta\mathcal{L} = \int_0^T \int_\Omega \lambda_1 \delta\rho\dot{v} - K^{-2}\delta K\lambda_2\dot{p} - \lambda_2\delta f(t)\delta(x - x_s)dxdt \tag{19}$$

Recalling that the first variation has the form of equation 6, we can write the gradients (functional derivatives) as follows

$$\frac{\delta\mathcal{L}}{\delta\rho(x)} = -\int_0^T \lambda_3(t)(x,t)\dot{v}(x,t)dt, \tag{20}$$

$$\frac{\delta\mathcal{L}}{\delta K(x)} = -\int_0^T \frac{1}{K^2}\lambda_2(x,t)\dot{p}(x,t)dt, \tag{21}$$

$$\frac{\delta\mathcal{L}}{\delta f(t)} = -\int_\Omega \lambda_2(x,t)\delta(x - x_s)dx. \tag{22}$$

Equations 20-22 are consistent with what is observed in the literature (Plessix, 2006) and amount to solving the forward PDE, calculating the residual, then injecting the residual at the receiver locations as a source to the adjoint equations and finally either cross-correlating the adjoint solution with a time-derivative of the forward solution (equations 20-21) or restricting the adjoint solution to the source location (equation 22).

ADOINT EQUATIONS AND GRADIENT FOR THE SEMI-DISCRETE PROBLEM

We now provide the derivation of the semi-discrete adjoint equation with SBP operators and the SAT method. We first begin by introducing SBP-SAT for both staggered and non-staggered grids. We then pose the optimization problem and derive the adjoint equations and gradient.

Summation-by-parts (SBP) finite-difference operators

SBP finite-difference operators are very useful in that they allow for the derivation of a numerical energy balance that closely resembles that of the continous energy balance. This in turn allows for explicit proofs of stability of the numerical scheme (Karlstrom and Dunham, 2016). In this section, we give basic definitions of SBP operators and summarize useful properties that we use in the derivations of the adjoint equations with staggered SBP spatial derivative operators.

For a semi-discrete PDE, a SBP spatial derivative has the form:

$$\mathbf{D} = \mathbf{H}^{-1}\mathbf{Q} \tag{23}$$

where $\mathbf{H}$ is a diagonal positive-definite matrix and $\mathbf{Q}$ is an almost skew-symmetric matrix such that $\mathbf{Q}^T + \mathbf{Q} = -\mathbf{E}_0 + \mathbf{E}_N$, where $\mathbf{E}_0$ and $\mathbf{E}_N$ are diagonal matrices with one non-zero element in either the first or last element of the diagonal respectively. As an example, the second-order accurate (on the interior) SBP operator is given by:

$$\mathbf{H} = h\,\mathbf{diag}\left[\begin{array}{ccccccc} \frac{1}{2} & 1 & 1 & \cdots & 1 & \frac{1}{2} \end{array}\right], \tag{24}$$

$$\mathbf{Q} = \frac{1}{2}\begin{bmatrix} -1 & 1 & & & & & \\ -1 & 0 & 1 & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & & -1 & 0 & 1 & \\ & & & & -1 & 1 \end{bmatrix}, \tag{25}$$

though of course higher-order methods can be utilized. Note also that the structure of $\mathbf{Q}$ will change depending on the treatment of the boundary conditions. For example, for periodic boundary conditions, $\mathbf{Q}$ is in fact skew-symmetric. Additionally, we can neglect to treat either the top or bottom boundary of the domain therefore leaving $\mathbf{Q}^T + \mathbf{Q} = -\mathbf{E}_0$.

The beauty of these operators is in the discrete analogy they make to integration-by-parts. First, notice that we can write:

$$\mathbf{HD} = -(\mathbf{HD})^T - \mathbf{E}_0 + \mathbf{E}_n \tag{26}$$

because $\mathbf{Q} = \mathbf{HD}$. Now if we define the inner-product between two continuous functions as:

$$(u, v) = \int_a^b u(x)v(x)dx$$

and therefore,

$$\left(u, \frac{dv}{dx}\right) = \int_a^b u(x)\frac{dv}{dx}dx = -\left(\frac{du}{dx}, v\right) + u(x)v(x)\Big|_a^b.$$

where we used integration by parts. In the discrete case, we can write the inner product of two vectors as:

$$(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T\mathbf{H}\mathbf{v}, \tag{27}$$

and now analogously, we can expand the inner product between a vector and the derivative of vector using SBP operators as:

$$\begin{aligned} (\mathbf{u}, \mathbf{Dv}) = \mathbf{u}^T\mathbf{HDv} &= \mathbf{u}^T(-(\mathbf{HD})^T - \mathbf{E}_0 + \mathbf{E}_n)\mathbf{v} \\ &= -\mathbf{u}^T(\mathbf{HD})^T\mathbf{v} - \mathbf{u}^T\mathbf{E}_0\mathbf{v} + \mathbf{u}^T\mathbf{E}_n\mathbf{v} \\ &= -\mathbf{u}^T\mathbf{D}^T\mathbf{H}\mathbf{v} - u_0v_0 + u_nv_n \\ &= -(\mathbf{Du}, \mathbf{v}) + u_nv_n - u_0v_0. \end{aligned}$$

As is clear in the above example, the summation-by-parts operators enable the switch of the derivative from $\mathbf{v}$ to $\mathbf{u}$ just as in integration-by-parts and additionally evaluate the difference of the product of the two functions at two points.

In addition to standard grids used in finite difference solutions to PDEs, the SBP property has also been extended to finite difference operators that operate on staggered grids. Following the work done by OReilly et al. (2016) the staggered SBP operator can be defined as

$$\mathbf{D}_{\pm} = \mathbf{H}_{\pm}^{-1}\mathbf{Q}_{\pm}$$

where $\pm$ indicates the fact that we now have two grids (the + grid and the - grid). These operators satisfy the following SBP property

$$\mathbf{H}_{+}\mathbf{D}_{+} + (\mathbf{H}_{-}\mathbf{D}_{-})^{T} = \mathbf{B}_{+} = \mathbf{B}_{-}^{T} = \mathbf{B}, \tag{28}$$

where the $\mathbf{B}$ matrix restricts the grid functions to the boundary

$$\boldsymbol{\phi}^{T}\mathbf{B}\boldsymbol{\psi} = \phi_N\psi_N - \phi_0\psi_0.$$

Again, these two expressions therefore allow us to mimic integration-by-parts in the following manner

$$\begin{aligned}(\boldsymbol{\phi}, \mathbf{D}_{+}\boldsymbol{\psi})_{H_{+}} = \boldsymbol{\phi}^{T}\mathbf{H}_{+}\mathbf{D}_{+}\boldsymbol{\psi} &= \boldsymbol{\phi}^{T}(-(\mathbf{H}_{-}\mathbf{D}_{-})^{T} + \mathbf{B})\boldsymbol{\psi} \\ &= -\boldsymbol{\phi}^{T}(\mathbf{H}_{-}\mathbf{D}_{-})^{T}\boldsymbol{\psi} + \phi_N\psi_N - \phi_0\psi_0 \\ &= -(\mathbf{D}_{-}\boldsymbol{\phi})^{T}\mathbf{H}_{-}\boldsymbol{\psi} + \phi_N\psi_N - \phi_0\psi_0 \\ &= -(\mathbf{D}_{-}\boldsymbol{\phi}, \boldsymbol{\psi})_{H_{-}} + \phi_N\psi_N - \phi_0.\psi_0\end{aligned}$$

Simultaneous approximation term (SAT) method

The SAT method used in conjunction with SBP operators for finite-difference methods, allow for weak enforcement of the boundary conditions (Karlstrom and Dunham, 2016). These terms are placed as penalty terms on the differential equation. As an example, consider equations 1-2 but without a source term. In the SBP-SAT framework, penalty terms change equations 1 and 2 in the following manner:

$$\rho\frac{\partial \mathbf{v}}{\partial t} = \cdots - c\mathbf{H}^{-1}\left[\mathbf{e}_0(v_0 - \hat{v}_0) + \mathbf{e}_n(v_n - \hat{v}_n)\right], \tag{29}$$

$$\frac{1}{K}\frac{\partial \mathbf{p}}{\partial t} = \cdots - c\mathbf{H}^{-1}\left[\mathbf{e}_0(p_0 - \hat{p}_0) + \mathbf{e}_n(p_n - \hat{p}_n)\right], \tag{30}$$

where c is a diagonal matrix containing values of the acoustic wave speed $c_i = \sqrt{\frac{K_i}{\rho_i}}$ and $\hat{v}_0, \hat{v}_n$ and $\hat{p}_n, \hat{p}_n$ are target values for the particle velocity and pressure at the first and last gridpoints respectively and are specified depending on the desired boundary condition.

In order to solve for the target values we first write out the following system of equations:

$$p_0 - \rho c v_0 = \hat{p}_0 - \rho c \hat{v}_0 \tag{31}$$
$$p_n - \rho c v_n = \hat{p}_n - \rho c \hat{v}_n, \tag{32}$$

which are the characteristics of the original PDE (equations 1-2). With equations 31-32 we can specify the desired boundary condition as a function of the target ("hat") variables and then solve for the penalty term. The inclusion of these penalty terms based on the different boundary conditions, will change the semi-discrete PDE and therefore how the spatial derivative operator and its adjoint are applied at the boundaries. Later in this report we provide an example of how this is done with the free-surface boundary condition.

Semi-discrete adjoint equations

In order to derive the adjoint equations and gradient for the SBP-SAT discretization, we first discretize the PDE in space so that it is semi-discrete. Then, we define the constrained optimization problem and derive the adjoint equations for periodic boundary conditions.

As we chose to perform the discretization with staggered grids, we have two different grids, which we denote as the (+) grid and the (-) grid. We define the solution vectors $\mathbf{v}$ and $\mathbf{p}$ on the (+) and (-) grids respectively. This then means that the material properties ρ and K will also be defined on the (+) and (-) grids respectively. We can explicitly write out the solution vectors and diagonal matrices containing the material properties as

$$\mathbf{v} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_N \end{bmatrix}, \quad \rho = \mathbf{diag}\left(\begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_N \end{bmatrix}\right)$$

$$\mathbf{p} = \begin{bmatrix} p_0 \\ p_{1/2} \\ \vdots \\ p_{N-1/2} \\ p_N \end{bmatrix}, \quad K = \mathbf{diag}\left(\begin{bmatrix} K_0 & K_{1/2} & \cdots & K_{n-1/2} & K_N \end{bmatrix}\right)$$

Figure 1 shows a pictorial description of the different grids. Note that with the staggered grids, the solution vectors (and therefore also the differential operators $\mathbf{D}_+$ and $\mathbf{D}_-$) no longer have the same dimension. Therefore, we find that for a grid of $N+1$ points we have that $\mathbf{D}_+ \in \mathbb{R}^{(N+1)\times(N+2)}$ and $\mathbf{H}_+ \in \mathbb{R}^{(N+1)\times(N+1)}$. Similarly, we find that for the (-) grid we have $\mathbf{D}_- \in \mathbb{R}^{(N+2)\times(N+1)}$ and $\mathbf{H}_- \in \mathbb{R}^{(N+2)\times(N+2)}$. Now we can write a semi-discrete version of the continuous forward problem without

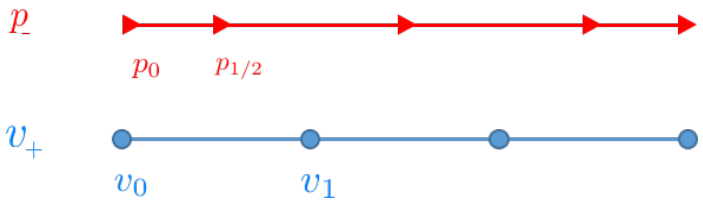

Figure 1: Solution vectors $\mathbf{p}$ and $\mathbf{v}$ positioned on staggered grids. Replicated after a similar figure in OReilly et al. (2016)

boundary terms (equations 1-2) as

$$\rho\dot{\mathbf{v}} + \mathbf{D}_+\mathbf{p} = \mathbf{0} \tag{33}$$
$$K^{-1}\dot{\mathbf{p}} + \mathbf{D}_-\mathbf{v} = \mathbf{d}_{s-} \tag{34}$$

where the discrete delta function $\mathbf{d}_{s-}$ is a high-order discretization of the delta function that restricts the source time function to the (-) grid (Petersson et al., 2016).

In order to setup the optimization problem, we first define the discrete parameters that we desire to estimate as vectors $\boldsymbol{\rho} \in \mathbb{R}^{N+1}$ and $\mathbf{K} \in \mathbb{R}^{N+2}$ (because the problem remains continuous in time, we still estimate a source time function $f(t)$). These are related to the diagonal matrices used in our semi-discrete PDE as $\rho = \text{diag}(\boldsymbol{\rho})$ and $K^{-1} = \text{diag}(\mathbf{K})^{-1}$. We now can write the optimization problem as

$$\begin{aligned}
&\underset{\boldsymbol{\rho},\mathbf{K},f(t)}{\text{minimize}} \quad F(\mathbf{v},\mathbf{p},\boldsymbol{\rho},\mathbf{K},f(t)), \quad \text{where} \\
&F(\mathbf{v},\mathbf{p},\boldsymbol{\rho},\mathbf{K},f(t)) = \frac{w_1}{2}\int_0^T \left((\mathbf{H}_+\mathbf{d}_{r+})^T\mathbf{v} - v_{\text{data}}(t)\right)^2 dt \\
&\qquad\qquad + \frac{w_2}{2}\int_0^T ((\mathbf{H}_-\mathbf{d}_{r-})^T\mathbf{p} - p_{\text{data}}(t))^2\, dt \\
&\text{subject to} \quad \rho\dot{\mathbf{v}} + \mathbf{D}_+\mathbf{p} = \mathbf{0}, \\
&\qquad\qquad K^{-1}\dot{\mathbf{p}} + \mathbf{D}_-\mathbf{v} - f(t)\mathbf{d}_{s-} = \mathbf{0},
\end{aligned} \tag{35}$$

where $\mathbf{d}_{r+}$ and $\mathbf{d}_{r-}$ are again high-order discretizations of the delta function at the receiver location and for the (+) and (-) grids respectively. As for the continuous problem, w_1 and w_2 are scalar weights that can be chosen depending on the data that we desire to fit. Now, introducing the semi-discrete Lagrange multipliers $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ we can write the unconstrained problem as

$$\begin{aligned}
\mathcal{L} = &\frac{w_1}{2}\int_0^T (\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t))^2 dt + \frac{w_2}{2}\int_0^T (\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t))^2 dt \\
&+ \int_0^T \boldsymbol{\lambda}_1^T\mathbf{H}_+\left(\rho\dot{\mathbf{v}} + \mathbf{D}_+\mathbf{p}\right) dt + \int_0^T \boldsymbol{\lambda}_2^T\mathbf{H}_-\left(K^{-1}\dot{\mathbf{p}} + \mathbf{D}_-\mathbf{v} - f(t)\mathbf{d}_{s-}\right) dt.
\end{aligned} \tag{36}$$

Note that we have introduced the Lagrange multipliers with the inner products $\boldsymbol{\lambda}_1\mathbf{H}_+$ and $\boldsymbol{\lambda}_2\mathbf{H}_-$. This is to remain consistent with our definition of the inner product between two vectors $(\mathbf{u},\mathbf{v})_H = \mathbf{u}^T\mathbf{H}\mathbf{v}$ where $\mathbf{H}$ is the quadrature operator as defined in equation 24. Again we calculate the first variation of the misfit functional. For the semi-discrete PDE problem the first variation will have the following generic form

$$\delta\mathcal{L} = \sum_{i=0}^{N}\frac{\partial\mathcal{L}}{\partial\rho_i}\delta\rho_i + \sum_{j=0}^{N+1}\frac{\partial\mathcal{L}}{\partial K_j}\delta K_j + \int_0^T \frac{\delta\mathcal{L}}{\delta f(t)}\delta f(t)dt. \tag{37}$$

We can then write the first variation of equation 36 as

$$\begin{aligned}
\delta\mathcal{L} = w_1 &\int_0^T \left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right)\mathbf{d}_{r+}^T\mathbf{H}_+\delta\mathbf{v}dt \\
&+ w_2\int_0^T \left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right)\mathbf{d}_{r-}^T\mathbf{H}_-\delta\mathbf{p}dt \\
&+ \int_0^T \boldsymbol{\lambda}_1^T\mathbf{H}_+\left(\left(\sum_{i=0}^{N}\frac{\partial\rho}{\partial\rho_i}\delta\rho_i\right)\dot{\mathbf{v}} + \rho\delta\dot{\mathbf{v}} + \mathbf{D}_+\delta\mathbf{p}\right)dt \\
&+ \int_0^T \boldsymbol{\lambda}_2^T\mathbf{H}_-\left(\left(\sum_{j=0}^{N+1}\frac{\partial K^{-1}}{\partial K_j}\delta K_j\right)\dot{\mathbf{p}} + K^{-1}\delta\dot{\mathbf{p}} + \mathbf{D}_-\delta\mathbf{v} - \delta f(t)\mathbf{d}_{s-}\right)dt
\end{aligned} \tag{38}$$

Note that because the diagonal matrices ρ and K^{-1} contain the same parameters as the vectors $\boldsymbol{\rho}$ and $\mathbf{K}$, we can simplify these sums as follows

$$\begin{aligned}
\delta\mathcal{L} = w_1 &\int_0^T \left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right)\mathbf{d}_{r+}^T\mathbf{H}_+\delta\mathbf{v}dt \\
&+ w_2\int_0^T \left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right)\mathbf{d}_{r-}^T\mathbf{H}_-\delta\mathbf{p}dt \\
&+ \int_0^T \boldsymbol{\lambda}_1^T\mathbf{H}_+\left(\delta\rho\dot{\mathbf{v}} + \rho\delta\dot{\mathbf{v}} + \mathbf{D}_+\delta\mathbf{p}\right)dt \\
&+ \int_0^T \boldsymbol{\lambda}_2^T\mathbf{H}_-\left(-K^{-2}\delta K\dot{\mathbf{p}} + K^{-1}\delta\dot{\mathbf{p}} + \mathbf{D}_-\delta\mathbf{v} - \delta f(t)\mathbf{d}_{s-}\right)dt,
\end{aligned} \tag{39}$$

where,

$$\delta\rho = \begin{bmatrix} \delta\rho_1 & & \\ & \ddots & \\ & & \delta\rho_{N+1} \end{bmatrix}, \tag{40}$$

$$-K^{-2}\delta K = -\begin{bmatrix} K_1^{-2} & & \\ & \ddots & \\ & & K_{N+2}^{-2} \end{bmatrix}\begin{bmatrix} \delta K_1 & & \\ & \ddots & \\ & & \delta K_{N+2} \end{bmatrix}. \tag{41}$$

Now as with the continuous case, we must integrate by parts in time to remove the time derivatives on $\delta\dot{\mathbf{p}}$ and $\delta\dot{\mathbf{v}}$. Doing so and combining all terms associated with the variations $\delta\mathbf{v}$ and $\delta\mathbf{p}$ we have the following

$$\begin{aligned}\delta\mathcal{L} = &\int_0^T \left(w_1\left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right)\mathbf{d}_{r+}^T\mathbf{H}_+ - \dot{\boldsymbol{\lambda}}_1^T\mathbf{H}_+\rho + \boldsymbol{\lambda}_2^T\mathbf{H}_-\mathbf{D}_-\right)\delta\mathbf{v}dt \\ &+ \int_0^T \left(w_2\left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right)\mathbf{d}_{r-}^T\mathbf{H}_- - \dot{\boldsymbol{\lambda}}_2^T\mathbf{H}_-\rho + \boldsymbol{\lambda}_1^T\mathbf{H}_+\mathbf{D}_+\right)\delta\mathbf{p}dt \\ &+ \boldsymbol{\lambda}_1(T)^T\mathbf{H}_+\delta\mathbf{v}(T) - \boldsymbol{\lambda}_1^T(0)\mathbf{H}_+\delta\mathbf{v}(0) + \boldsymbol{\lambda}_2(T)^T\mathbf{H}_-\delta\mathbf{p}(T) - \boldsymbol{\lambda}_2^T(0)\mathbf{H}_-\delta\mathbf{p}(0) \\ &+ \int_0^T \boldsymbol{\lambda}_1^T\mathbf{H}_+\delta\rho\dot{\mathbf{v}} - \boldsymbol{\lambda}_2^T\mathbf{H}_-K^{-2}\delta K\dot{\mathbf{p}} - \boldsymbol{\lambda}_2^T\mathbf{H}_-\mathbf{d}_{s-}\delta f(t)dt. \end{aligned}\tag{42}$$

Again, we impose conditions on the Lagrange multipliers to avoid calculating $\delta\mathbf{p}$ and $\delta\mathbf{v}$. This results in the following adjoint equations with terminal conditions

$$\dot{\boldsymbol{\lambda}}_1^T\mathbf{H}_+\rho - \boldsymbol{\lambda}_2^T\mathbf{H}_-\mathbf{D}_- = w_1\left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right)\mathbf{d}_{r+}^T\mathbf{H}_+, \tag{43}$$

$$\dot{\boldsymbol{\lambda}}_2^T\mathbf{H}_-K^{-1} - \boldsymbol{\lambda}_1^T\mathbf{H}_+\mathbf{D}_+ = w_2\left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right)\mathbf{d}_{r-}^T\mathbf{H}_-, \tag{44}$$

$$\boldsymbol{\lambda}_1(T) = \mathbf{0}, \quad \boldsymbol{\lambda}_2(T) = \mathbf{0}. \tag{45}$$

Again, we use the fact that $\mathbf{v}(0) \Rightarrow \delta\mathbf{v}(0)$ and $\mathbf{p}(0) \Rightarrow \delta\mathbf{p}(0)$ which eliminates the need for imposing conditions on $\boldsymbol{\lambda}_1(0)$ and $\boldsymbol{\lambda}_2(0)$. Now we transpose both adjoint equations and use the staggered SBP property $(\mathbf{D}_-\mathbf{H}_-)^T = -\mathbf{H}_+\mathbf{D}_+$ (equation 28) to obtain the following equations

$$\mathbf{H}_+\rho\dot{\boldsymbol{\lambda}}_1 + \mathbf{H}_+\mathbf{D}_+\boldsymbol{\lambda}_2 = w_1\mathbf{H}_+\mathbf{d}_{r+}\left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right), \tag{46}$$

$$\mathbf{H}_-K^{-1}\dot{\boldsymbol{\lambda}}_2 + \mathbf{H}_-\mathbf{D}_-\boldsymbol{\lambda}_1 = w_2\mathbf{H}_-\mathbf{d}_{r-}\left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right). \tag{47}$$

Multiplying equation 46 through by $\mathbf{H}_+^{-1}$ and equation 47 by $\mathbf{H}_-^{-1}$ we obtain

$$\rho\dot{\boldsymbol{\lambda}}_1 + \mathbf{D}_+\boldsymbol{\lambda}_2 = w_1\mathbf{d}_{r+}\left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right), \tag{48}$$

$$K^{-1}\dot{\boldsymbol{\lambda}}_2 + \mathbf{D}_-\boldsymbol{\lambda}_1 = w_2\mathbf{d}_{r-}\left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right). \tag{49}$$

Again, we change temporal coordinates $\tau = T - t$ and let $\boldsymbol{\lambda}_3 = -\boldsymbol{\lambda}_1$ which results in the following self-adjoint equations

$$\rho\boldsymbol{\lambda}_3' + \mathbf{D}_+\boldsymbol{\lambda}_2 = w_1\mathbf{d}_{r+}\left(\mathbf{d}_{r+}^T\mathbf{H}_+\mathbf{v} - v_{\text{data}}(t)\right), \tag{50}$$

$$K^{-1}\boldsymbol{\lambda}_2' + \mathbf{D}_-\boldsymbol{\lambda}_3 = -w_2\mathbf{d}_{r-}\left(\mathbf{d}_{r-}^T\mathbf{H}_-\mathbf{p} - p_{\text{data}}(t)\right), \tag{51}$$

$$\boldsymbol{\lambda}_1(\tau = 0) = \mathbf{0}, \quad \boldsymbol{\lambda}_2(\tau = 0) = \mathbf{0}. \tag{52}$$

Upon satisfying equations 50-52, our expression for the first variation becomes

$$\begin{aligned}\delta\mathcal{L} &= \int_0^T \boldsymbol{\lambda}_1^T \mathbf{H}_+ \delta\rho\dot{\mathbf{v}} - \boldsymbol{\lambda}_2^T \mathbf{H}_- K^{-2}\delta K\dot{\mathbf{p}} - \boldsymbol{\lambda}_2^T \mathbf{H}_- \mathbf{d}_{s-}\delta f(t)dt \\ &= \int_0^T -\sum_{i=0}^{N} \lambda_{3i} H_{+ii}\dot{v}_i\delta\rho_i - \sum_{j=0}^{N+1}\frac{1}{K^2}\lambda_{2j}H_{-jj}\dot{p}_j\delta K_j + \boldsymbol{\lambda}_2^T\mathbf{H}_-\mathbf{d}_{s-}\delta f(t)dt. \end{aligned} \tag{53}$$

We can now write the expressions for the gradients for our parameters $\boldsymbol{\rho}$, $\mathbf{K}$ and $f(t)$ as

$$\frac{\partial\mathcal{L}}{\partial\rho_i} = -\int_0^T \lambda_{3i}(t)\dot{v}(t)_i H_{+ii}dt, \tag{54}$$

$$\frac{\partial\mathcal{L}}{\partial K_j} = -\int_0^T \frac{1}{K_j^2}\lambda_2(t)\dot{p}_j(t)H_{-jj}dt, \tag{55}$$

$$\frac{\delta\mathcal{L}}{\delta f(t)} = -\boldsymbol{\lambda}_2^T(t)\mathbf{H}_-\mathbf{d}_{s-} \tag{56}$$

where as for the continuous case, we compared equation 53 with equation 37 in order to determine the expression of the gradients. Comparing equations 54-56 with equations 20-22 we observe that our gradients for the semi-discrete problem are consistent with the continuous problem. We also observe that our semi-discrete adjoint equations (50-51) are consistent with the continuous adjoint equations (16-17).

Adjoint SBP operator

If we combine $\mathbf{v}$ and $\mathbf{p}$ into a single composite vector $\mathbf{q}$ we can write a single composite semi-discrete PDE as

$$\mathbf{C}\dot{\mathbf{q}} - \mathbf{A}\mathbf{q} = \mathbf{d}_s f(t), \tag{57}$$

where

$$\mathbf{q} = \begin{bmatrix}\mathbf{v}\\ \mathbf{p}\end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix}\mathbf{0} & -\mathbf{D}_+\\ -\mathbf{D}_- & \mathbf{0}\end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix}\rho & \mathbf{0}\\ \mathbf{0} & K^{-1}\end{bmatrix}, \quad \mathbf{d}_s = \begin{bmatrix}\mathbf{0}\\ \mathbf{d}_{s-}\end{bmatrix}. \tag{58}$$

The adjoint equations (equations 50-51) can then be written in the following composite form

$$\mathbf{C}\boldsymbol{\lambda}' - \mathbf{H}^{-1}\mathbf{A}^T\mathbf{H}\boldsymbol{\lambda} = \mathbf{d}_r\mathbf{a}(t), \tag{59}$$

where

$$\boldsymbol{\lambda} = \begin{bmatrix}\boldsymbol{\lambda}_3\\ \boldsymbol{\lambda}_2\end{bmatrix}, \mathbf{H} = \begin{bmatrix}\mathbf{H}_+ & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_-\end{bmatrix}, \mathbf{d}_r = \begin{bmatrix}\mathbf{d}_{r+} & 0\\ 0 & \mathbf{d}_{r-}\end{bmatrix}, \tag{60}$$

$$\mathbf{a}(t) = \begin{bmatrix} w_1 \left(\mathbf{d}_{r+}^T \mathbf{H}_+ \mathbf{v} - v_{\text{data}}(t) \right) \\ -w_2 \left(\mathbf{d}_{r-}^T \mathbf{H}_- \mathbf{p} - p_{\text{data}}(t) \right) \end{bmatrix}. \tag{61}$$

Comparing equations 57 and 58, we observe that equation 58 differs in that it is solved in a reverse time coordinate system with an adjoint spatial derivative operator $\mathbf{A}^\dagger = \mathbf{H}^{-1}\mathbf{A}^T\mathbf{H}$ and with an adjoint source $\mathbf{a}(t)$ that is injected at the receiver location. We point out that the adjoint operator $\mathbf{A}^\dagger$ is not simply $\mathbf{A}^T$ which is what is commonly found when deriving adjoint equations via the adjoint method. The reason for this is because we have defined our inner product as defined in equation 27 as opposed to the standard L2 inner product. This therefore leads to a different norm and different way of defining the adjoint operator.

In fact, here we show that $\mathbf{A}^\dagger = \mathbf{H}^{-1}\mathbf{A}^T\mathbf{H}$ is skew self-adjoint. We can check for the adjointness of this operator by simply expanding $\mathbf{H}^{-1}\mathbf{A}^T\mathbf{H}$ as follows $\mathbf{H}^{-1}\mathbf{A}^T\mathbf{H}$

$$\begin{aligned}
\mathbf{H}^{-1}\mathbf{A}^T\mathbf{H} &= \begin{bmatrix} \mathbf{H}_+^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_-^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & -\mathbf{D}_-^T \\ -\mathbf{D}_+^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{H}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_- \end{bmatrix}, \\
&= \begin{bmatrix} \mathbf{H}_+^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_-^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & -(\mathbf{H}_-\mathbf{D}_-)^T \\ -(\mathbf{H}_+\mathbf{D}_+)^T & \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{H}_+^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_-^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{H}_+\mathbf{D}_+ \\ \mathbf{H}_-\mathbf{D}_- & \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0} & \mathbf{D}_+ \\ \mathbf{D}_- & \mathbf{0} \end{bmatrix} = -\mathbf{A}
\end{aligned}$$

where again we used the staggered SBP property $(\mathbf{D}_-\mathbf{H}_-)^T = -\mathbf{H}_+\mathbf{D}_+$. Thus the staggered adjoint operator $\mathbf{A}^\dagger = \mathbf{H}^{-1}\mathbf{A}^T\mathbf{H}$ is skew self-adjoint.

Example with a free-surface boundary condition

Finally, we demonstrate how to write out the forward problem for the free-surface boundary condition at the 0th gridpoint and obtain the forward operator $\mathbf{A}$. As stated before, we start from the incoming/outgoing characteristic of the PDE at the left boundary (0th gridpoint)

$$p_0 - \rho c v_0 = \hat{p}_0 - \rho c \hat{v}_0. \tag{62}$$

To enforce the free-surface boundary condition, we set the target variable $\hat{p}_0 = 0 \Rightarrow v_0 - \hat{v}_0 = \rho c(v_0 - \hat{v}_0)$ and then substitute these expressions into equations 29 and 30. Then then gives us the following semidiscrete ODE

$$\rho \dot{\mathbf{v}} + \mathbf{D}_+\mathbf{p} - \mathbf{H}_+^{-1}\mathbf{B}_+\mathbf{p} = \mathbf{0}, \tag{63}$$

$$K^{-1}\dot{\mathbf{p}} + \mathbf{D}_-\mathbf{v} + c_0 K^{-1}\mathbf{H}_-^{-1}\mathbf{E}_{0-}\mathbf{p} = \mathbf{0}. \tag{64}$$

Writing equations 63 and 64 in block form, we have the following expression

$$\begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & K^{-1} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{v}} \\ \dot{\mathbf{p}} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & -\mathbf{D}_+ + \mathbf{H}_+^{-1}\mathbf{B}_+ \\ -\mathbf{D}_- & -c_0 K^{-1}\mathbf{H}_-^{-1}\mathbf{E}_{0-} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \end{bmatrix} = \mathbf{d}f(t). \tag{65}$$

Note now that our spatial derivative operator $\mathbf{A}$ no longer has zeros on the diagonal and also has an additional term added to the derivative operator. By definition, the adjoint operator is

$$\mathbf{A}^\dagger = \begin{bmatrix} \mathbf{H}_+^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_-^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & -\mathbf{D}_-^T \\ -\mathbf{D}_+^T + \mathbf{B}_+^T\mathbf{H}_+^{-1} & -c_0K^{-1}\mathbf{H}_-^{-1}\mathbf{E}_{0-} \end{bmatrix} \begin{bmatrix} \mathbf{H}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_- \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}_+^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_-^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & -\mathbf{D}_-^T\mathbf{H}_-^{-1} \\ -\mathbf{D}_+^T\mathbf{H}_+ + \mathbf{B}_+^T & -c_0K^{-1}\mathbf{E}_{0-} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}_+^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_-^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{H}_+\mathbf{D}_+ - \mathbf{B}_+ \\ \mathbf{H}_-\mathbf{D}_- & -c_0K^{-1}\mathbf{E}_{0-} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{D}_+ - \mathbf{H}_+^{-1}\mathbf{B}_+ \\ \mathbf{D}_- & -c_0K^{-1}\mathbf{H}^{-1}\mathbf{E}_{0-} \end{bmatrix}$$

With this adjoint operator, we obtain the following adjoint equations

$$\begin{aligned} \rho\boldsymbol{\lambda}_3' + \mathbf{D}_+\boldsymbol{\lambda}_2 - \mathbf{H}_+^{-1}\mathbf{B}_+\boldsymbol{\lambda}_2 &= \mathbf{0}, \\ K^{-1}\boldsymbol{\lambda}_2' + \mathbf{D}_-\boldsymbol{\lambda}_3 + c_0K^{-1}\mathbf{H}_-^{-1}\mathbf{E}_{0-}\boldsymbol{\lambda}_2 &= \mathbf{0}, \end{aligned}$$

which are a consistent approximation of the adjoint equations with the dual boundary condition $\lambda_2(0,t) = 0$.

NUMERICAL EXAMPLE

We verified our derived adjoint equations and gradient with a numerical example in which we estimate a source-time function $f(t)$ from a synthetic pressure time series. The true data (p_{data}) were modeled with a Gaussian wavelet injected with our discrete form of the delta function $\mathbf{d}_{s-}$. This Gaussian wavelet is shown in Figure 2a and is the wavelet we desire to estimate from the modeled data. The time-stepper we used in the modeling was a fourth-order Runge-Kutta stepper. For the boundary conditions, we used absorbing boundary conditions. Figure 2b shows the wavelet used for the initial source in the inversion which was a time-shifted version of the true wavelet. Figure 2c shows the adjoint source input to the adjoint equations (the right hand side of equation 51). While not shown in Figure 2, solving the adjoint equations with the adjoint source will result in a wavefield that will effectively cancel the location of the initial source and put a source at the location of the true source. The estimated source wavelet after running a steepest-descent optimization is shown in Figure 2d.

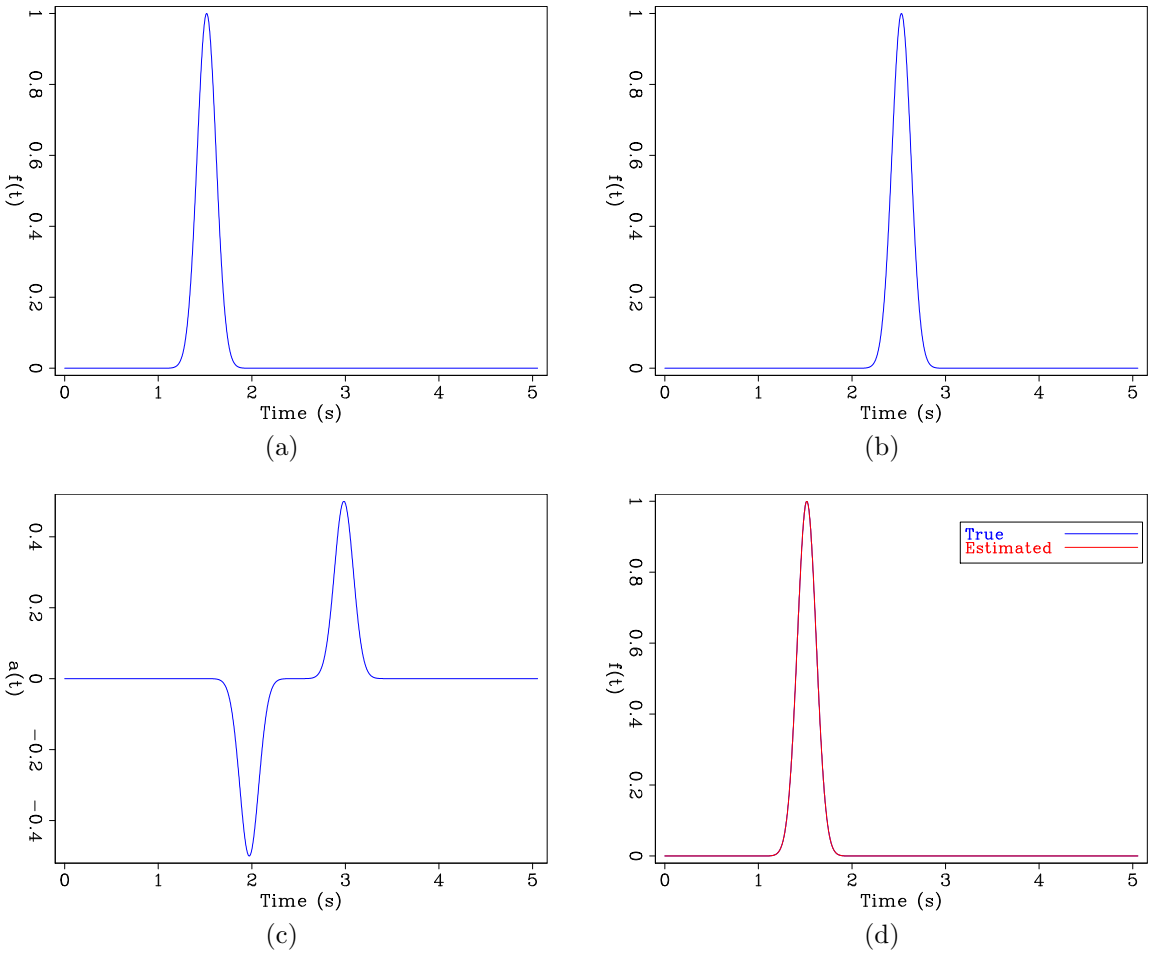

Figure 2: The results of the numerical test. (a) The true source wavelet we desire to estimate. (b) The initial source used in the optimization which is just a time-shifted version of the true source. (c) The adjoint source that is input to the adjoint equations. (d) Estimated Gaussian wavlet. [**CR**]

DISCUSSION AND CONCLUSION

We applied the theory of high-order finite-difference operators that satisfy the summation-by-parts property to the FWI problem (a PDE-constrained optimization problem). We used the adjoint method to find gradient of this optimization problem which requires first solving a system of adjoint equations and then with the solution of the adjoint solution the gradient can be calculated. The derivation of the semi-discrete adjoint equations using SBP operators is mathematically an intuitive extension of the continuous derivation. For the continuous problem, an integration by parts in space and in time is required in order to derive the adjoint equations. For the semi-discrete problem, an integration by parts in time and a summation-by-parts in space was required to derive the adjoint equations. We showed for our example that these SBP operators lead to self-adjoint spatial discretizations. Therefore, only one code was required to solve both forward and adjoint equations. Additionally, we showed how to derive the adjoint operator for a free-surface boundary condition. For a simple numerical example we estimated a source wavelet from synthetic pressure data. In the future, we plan to provide numerical examples estimating the medium parameters ρ and K and additionally show examples of modeling the adjoint source at a fluid-solid interface that must be done for elastic FWI for ocean-bottom node data.

REFERENCES

Alves, G. C., 2017, Elastic full waveform inversion of multicomponent data: PhD thesis, Stanford University.

Berg, J. and J. Nordström, 2012, Superconvergent functional output for time-dependent problems using finite differences on summation-by-parts form: Journal of Computational Physics, **231**, 6846–6860.

Biondi, E., B. Biondi, and R. Clapp, 2016, Wavefield-based AVO Inversion: Elastic images from pressure waves: SEP-Report, **163**, 33–40.

Bradley, A. M., 2013, Pde-constrained optimization and the adjoint method: `http://cs.stanford.edu/~ambrad/adjoint_tutorial.pdf`.

Duru, K. and E. M. Dunham, 2016, Dynamic earthquake rupture simulations on non-planar faults embedded in 3d geometrically complex, heterogeneous elastic solids: Journal of Computational Physics, **305**, 185–207.

Karlstrom, L. and E. M. Dunham, 2016, Excitation and resonance of acoustic-gravity waves in a column of stratified, bubbly magma: Journal of Fluid Mechanics, **797**, 431–470.

Liu, Q. and J. Tromp, 2006, Finite-frequency kernels based on adjoint methods: Bulletin of the Seismological Society of America, **96**, 2383–2397.

———, 2008, Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods: Geophysical Journal International, **174**, 265–286.

Lotto, G. C. and E. M. Dunham, 2015, High-order finite difference modeling of tsunami generation in a compressible ocean from offshore earthquakes: Computational Geosciences, **19**, 327–340.

Ober, C., T. Smith, J. Overfelt, S. Collis, G. Von Winckel, B. Van Bloemen Waanders, N. Downey, S. Mitchell, S. Bond, D. Aldridge, et al., 2016, Visco-tti-elastic fwi using discontinuous galerkin: Presented at the 2016 SEG International Exposition and Annual Meeting.

OReilly, O., T. Lundquist, J. Nordström, and E. M. Dunham, 2016, Energy stable and high-order-accurate finite difference methods on staggered grids.

Petersson, N. A., O. O'Reilly, B. Sjögreen, and S. Bydlon, 2016, Discretizing singular point sources in hyperbolic wave propagation problems: Journal of Computational Physics, **321**, 532–555.

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

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 2: Imaging compressive-wave and shear-wave velocities: Geophysical Journal International, **194**, 1665–1681.

Sears, T. J., P. J. Barton, and S. C. Singh, 2010, Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to alba field, uk north sea: Geophysics, **75**, R109–R119.

Sjögreen, B. and N. A. Petersson, 2014, Source estimation by full wave form inversion: Journal of Scientific Computing, **59**, 247–276.

Wang, Y., H. Zhou, S. Yuan, and Y. Ye, 2017, A fourth order accuracy summation-by-parts finite difference scheme for acoustic reverse time migration in boundary-conforming grids: Journal of Applied Geophysics, **136**, 498–512.