Gradient of image-space wave-equation tomography in the generalized source domain by the adjoint-state method

Claudio Guerra

Stanford Exploration Project, Mitchell Bldg., Department of Geophysics,

Stanford University, Stanford, CA 94305-2215

(September 24, 2009)

ABSTRACT

Optimization using gradient descent techniques requires the computation of the gradient of the objective function. The adjoint-state method is an efficient way of computing the gradient, since is not necessary to compute the expensive Frechét derivatives. Here, I derive the gradient of the image-space wave-equation tomography using the adjoint state method and show its application with a numerical example.

INTRODUCTION

Wave-equation tomography aims to solve for Earth models that explain the data under some norm. In general, two categories are distinguished depending on the domain the objective function is computed. In one category, known as waveform inversion (**???**), the objective function is defined in the data space where the modeled data are compared with the recorded seismograms. In the other category, generally called here as image-space wave-equation tomography (ISWET) from which wave-equation migration velocity analysis (WEMVA) (**??**) and differential semblance velocity analysis (DSVA) (**??**) are two variants, the objective function to be minimized is defined in the image space.

Despite their differences, those two categories are linked by the concept of extended modeling (?). ? showed that ISWET can be regarded as a solution method for the partially linearized waveform inversion. While ISWET solves for the long-wavelength components of the slowness model, assuring global convergence, waveform inversion computes the short-wavelength components.

Similar to waveform inversion, ISWET is a computationally demanding process. To decrease its computational cost it is common practice to use generalized sources (??). Because wavefield propagation is a linear process, generalized sources can be computed by linearly combining source wavefields and receiver wavefields, using phase encoding techniques (??). In particular, ? used generalized source functions synthesized by phase encoding modeling experiments in the image space drastically decreasing the cost of DSVA. The example showed in this report also uses this kind of generalized source functions.

The aforementioned ISWET methods differ on the way the objective function is computed and, consequently, the numerical optimization scheme applied. As pointed out by ?, in WEMVA, an intermediate residual-moveout parameter picking is necessary to produce the perturbed image via linearized residual prestack depth migration (?). The picking procedure makes it difficult to automate the whole process. As the computed perturbed image is consistent with the wave-equation tomographic operator, however, conjugate gradient methods can be used to invert for the slowness perturbation. In DSVA, the perturbed image is computed by applying the fully automated differential-semblance operator (DSO) on the background image, both using the subsurface-offset gathers (ODCIG) or angle gathers (ADCIG). When applied to ODCIG, DSVA minimizes the energy not focused at zero-offset. When applied to ADCIG, DSVA minimizes energy departing from flatness of the reflectors. Notwithstanding the ability of automating ISWET, DSVA produces perturbed images not consistent with the wave-equation tomographic operator. To minimize the objective function computed with DSO, quasi-Newton algorithms can be used for which the gradient of the objective function needs to be computed.

The gradient of the objective function can be obtained by computing the Frechét derivatives. However, this computation even for 2D applications of ISWET can be very expensive. An efficient way of computing the gradient without the need of Frechét derivatives is the adjoint-state method (**??**). **?** describes two methodologies for computing the gradient of the objective function using the adjoint-state method. The more intuitive is the one that uses the augmented Lagrangian. The augmented Lagrangian is formed by the objective function and the scalar product of the adjoint-state variables and general solutions of the forward modeling equations. The adjoint-state variables are solution of a system of adjoint-state equations defined by equating to zero the derivative of the augmented Lagrangian with respect to the state variables. For the linear case, the adjoint of the modeling operator applied to the adjoint-state variables gives the gradient of the objective function.

Here, I show how to compute the gradient of the ISWET objective function using the augmented Lagrangian methodology. The derivation of the gradient presented here is valid whether ISWET uses areal-shot migration or shot-profile migration. **?** provide a complete derivation of the forward and adjoint wave-equation tomographic operators in the generalized source domain. The goal of this paper is to provide a more detailed derivation of the gradient of the ISWET objective function using the adjoint state method than the ones available, for example, **?**.

The references on the image-space phase-encoded gathers are **??** and **?**. For completeness, here I briefly describe how to compute these phase-encoded gathers. Then, I derive the ISWET gradient using the adjoint-state method and, finally show an example of DSVA using image-space phase-encoded gathers on the Marmousi model.

IMAGE-SPACE GENERALIZED SOURCES

The prestack exploding-reflector modeling, introduced by **?** and **?**, synthesizes areal data and the corresponding areal source function, having as initial condition, a wave-equation migrated prestack image obtained with an inaccurate slowness model. In general, one single reflection event from one single ODCIG is modeled by the recursive upward continuation with the following one-way wave equations:

$$\begin{cases} \left(\frac{\partial}{\partial z} - i\sqrt{\omega^2 \hat{s}^2(\mathbf{x}) - |\mathbf{k}|^2}\right) d(\mathbf{x},\omega;x_m,y_m) = r_D(\mathbf{x},\mathbf{h};x_m,y_m) \\ d(x,y,z=z_{\max},\omega;x_m,y_m) = 0 \end{cases}, \tag{1}$$

and

$$\begin{cases} \left(\frac{\partial}{\partial z} + i\sqrt{\omega^2 \hat{s}^2(\mathbf{x}) - |\mathbf{k}|^2}\right) u(\mathbf{x},\omega;x_m,y_m) = r_U(\mathbf{x},\mathbf{h};x_m,y_m) \\ u(x,y,z=z_{\max},\omega;x_m,y_m) = 0 \end{cases}, \tag{2}$$

where $r_D(\mathbf{x},\mathbf{h};x_m,y_m)$ and $r_U(\mathbf{x},\mathbf{h};x_m,y_m)$ are the isolated SODCIGs at the horizontal location (x_m,y_m) for a single reflector, and are suitable for the initial conditions for the source and receiver wavefields, respectively. They are obtained by rotating the original unfocused SODCIGs according to the apparent geological dip of the reflector. This rotation maintains the velocity information needed for migration velocity analysis, especially for dipping reflectors (**?**). $d(x,y,z=z_c,\omega;x_m,y_m)$ is the areal source data and $u(x,y,z=z_c,\omega;x_m,y_m)$ is the areal receiver data for a single reflector and a single SODCIG located at (x_m,y_m). $z=z_c$ denotes that the wavefields can be collected at any depth level, z_c. This characteristic is important to accelerate ISWET especially if z_c is such that it separates regions of sufficiently accurate slowness above and inaccurate slowness below. As a consequence, the synthesized gathers are naturally 'datumized' and the wavefield propagations during ISWET can be restricted to the region where the slowness model must be updated. This feature allows ISWET be applied in a target-oriented manner.

However, if one considers to model one single reflection from one single SODCIG at a time, the modeled dataset can be orders of magnitude bigger than the original data set. As discussed by **???**, by modeling several reflectors and several SODCIGs simultaneously using random phase encoding a much smaller dataset is obtained. The randomly encoded areal source and areal receiver wavefields can be computed as follows:

$$\begin{cases} \left(\frac{\partial}{\partial z} - i\sqrt{\omega^2 \hat{s}^2(\mathbf{x}) - |\mathbf{k}|^2}\right) \widetilde{d}(\mathbf{x}, \mathbf{p}_m, \omega) = \widetilde{r}_D(\mathbf{x}, \mathbf{h}, \mathbf{p}_m, \omega) \\ \widetilde{d}(x, y, z = z_{\max}, \mathbf{p}_m, \omega) = 0 \end{cases}, \tag{3}$$

and

$$\begin{cases} \left(\frac{\partial}{\partial z} + i\sqrt{\omega^2 \hat{s}^2(\mathbf{x}) - |\mathbf{k}|^2}\right) \widetilde{u}(\mathbf{x}, \mathbf{p}_m, \omega) = \widetilde{r}_U(\mathbf{x}, \mathbf{h}, \mathbf{p}_m, \omega) \\ \widetilde{u}(x, y, z = z_{\max}, \mathbf{p}_m, \omega) = 0 \end{cases}, \tag{4}$$

where $\widetilde{r}_D(\mathbf{x}, \mathbf{h}, \mathbf{p}_m, \omega)$ and $\widetilde{r}_U(\mathbf{x}, \mathbf{h}, \mathbf{p}_m, \omega)$ are the encoded SODCIGs after rotations. They are defined as follows:

$$\widetilde{r}_D(\mathbf{x}, \mathbf{h}, \mathbf{p}_m, \omega) = \sum_{x_m} \sum_{y_m} r_D(\mathbf{x}, \mathbf{h}, x_m, y_m) \beta(\mathbf{x}, x_m, y_m, \mathbf{p}_m, \omega), \tag{5}$$

$$\widetilde{r}_U(\mathbf{x}, \mathbf{h}, \mathbf{p}_m, \omega) = \sum_{x_m} \sum_{y_m} r_U(\mathbf{x}, \mathbf{h}, x_m, y_m) \beta(\mathbf{x}, x_m, y_m, \mathbf{p}_m, \omega), \tag{6}$$

where $\beta(\mathbf{x}, x_m, y_m, \mathbf{p}_m, \omega) = e^{i\gamma(\mathbf{x}, x_m, y_m, \mathbf{p}_m, \omega)}$ is chosen to be the random phase-encoding function, with $\gamma(\mathbf{x}, x_m, y_m, \mathbf{p}_m, \omega)$ being a uniformly distributed random sequence in $\mathbf{x}$, x_m, y_m and ω; the variable $\mathbf{p}_m$ is the index of different realizations of the random sequence. Recursively solving Equations **??** and **??** gives us the encoded areal source data $\widetilde{d}(x, y, z = z_c, \mathbf{p}_m, \omega)$ and areal receiver data $\widetilde{u}(x, y, z = z_c, \mathbf{p}_m, \omega)$, which can be collected at the depth level, z_c.

GRADIENT COMPUTATION OF ISWET BY THE ADJOINT STATE METHOD

Image-space wave-equation tomography aims to solve for the slowness model, $s = s(\mathbf{x})$, that minimizes the linearized objective function

$$J(s) = \frac{1}{2} \left|\left|\Delta r(s)\right|\right|^2 = \frac{1}{2} \left|\left|r(s) - \mathbf{M} r(s)\right|\right|^2, \tag{7}$$

where $\Delta r = \Delta r(\mathbf{x}, \mathbf{h})$ is the image perturbation which measures the goodness of the slowness model. Notice that the objective function depends on the slowness model through the image perturbation. Δr is computed by applying a differential residual-focusing operator $\mathbf{M}$ (**?**), using either differential residual prestack-migration (**??**) or differential-semblance optimization (DSO) (**?**) operators to the image $r = r(\mathbf{x}, \mathbf{h})$. Herein, operators are represented by bold capital letters.

If the differential residual-focusing operator $\mathbf{M}$ is independent on the slowness, the gradient of this objective function, evaluated at the current slowness, $\widehat{s} = \widehat{s}(\mathbf{x})$, is

$$\nabla J(s) = \left(\frac{\partial r}{\partial s}\right)'\Bigg|_{s=\widehat{s}} \left(\mathbf{I} - \mathbf{M}'\right) \Delta \widehat{r}. \tag{8}$$

where '′' denotes the adjoint, $\mathbf{I}$ is the identity operator and $\Delta \widehat{r} = \Delta \widehat{r}(\mathbf{x}, \mathbf{h})$ is the perturbed image obtained with the current slowness model. The linear operator $\frac{\partial \widehat{r}}{\partial s}$ defines the mapping, $\frac{\partial \widehat{r}}{\partial s} \Delta s = \Delta r$, between the slowness perturbation Δs and the image perturbation and it is called image-space wave-equation tomographic operator.

As the image-space wave-equation tomographic operator is composed of different operators, it is difficult from equation **??** to envision which operations are performed to compute the gradient. Therefore, for a clear explanation of the operators involved, I use the adjoint-state method to derive the gradient of the objective function (equation **??**).

In migration with generalized sources (areal shot and shot-profile migration, e.g.), the source and receiver wavefields are propagated independently and the image, $r_z = r_z(\mathbf{x}, \mathbf{h})$, at a depth level z, is computed by the crosscorrelation

$$r_z(\mathbf{x},\mathbf{h}) = \sum_{\omega} d_z^*(\mathbf{x}-\mathbf{h},\omega) u_z(\mathbf{x}+\mathbf{h},\omega), \tag{9}$$

where $d_z(\mathbf{x},\omega)$ is the source wavefield for a single frequency ω at horizontal coordinates $\mathbf{x} = (x,y)$; $u_z(\mathbf{x},\omega)$ is the receiver wavefield and $\mathbf{h} = (h_x,h_y)$ is the subsurface half-offset, and '*' stands for the complex-conjugate. An additional summation over shots is required when migrating more than one shot. Hereafter, letters d and u stand for source and receiver wavefields, respectively, irrespective to the migration scheme.

In a more compact notation, not explicitly writing the dependencies on $\mathbf{x}$ and $\mathbf{h}$, equation **??** can be written as:

$$r_z = \mathbf{S}\mathbf{D}_z'(\omega) u_z(\omega) = \mathbf{S}\mathbf{U}_z(\omega) d_z^*(\omega), \tag{10}$$

where $\mathbf{D}$ and $\mathbf{U}$ are convolutional operators composed of (h_x,h_y)-shifted versions of $d_z(\mathbf{x},\omega)$ and $u_z(\mathbf{x},\omega)$, respectively. Operator $\mathbf{S}$ corresponds to the summation over frequency.

For subsequent depth levels, $d(\mathbf{x},\omega)$ is computed by means of the recursive downward propagation

$$\begin{cases} d_{z+1}(\omega) = \mathbf{T}_z^{\downarrow}(\omega,s) d_z(\omega) \\ d_1(\omega) = q(\omega), \end{cases} \tag{11}$$

where $\mathbf{T}_z^{\downarrow}$ is the downward continuation operator, which is function of the slowness, s, and $q(\omega)$ is the source wavefield used as boundary condition. In the case of conventional shot-profile migration $q(\omega) = f_s(\omega)\delta(\mathbf{x}-\mathbf{x_s})$ is the source signature located at $\mathbf{x}_s = (x_s,y_s,0)$. If using the generalized sources in the image space, $q(\omega)$ represents the image-space phase-encoded source wavefield of equation **??**.

The downward continuation of the receiver wavefield is performed by

$$\begin{cases} u_{z+1}(\omega) = \mathbf{T}_z^{\downarrow}(\omega, s)u_z(\omega) \\ u_1(\omega) = w(\omega), \end{cases} \tag{12}$$

where $w(\omega)$ is the recorded data at the surface for shot-profile migration. If using generalized sources in the image space, $w(\omega)$ is the phase-encoded areal receiver wavefield of equation**??**. In equations **??** and **??**, I omitted the dependencies of the wavefield with respect to $\mathbf{x}$. The subscript 1 in equations **??** and **??** represents the surface for the shot-profile migration and the 'collection' depth level, z_c, for the image-space phase-encoded wavefields.

In the image-space wave-equation tomography problem, the perturbed source and receiver wavefields and image perturbations are used to compute the slowness perturbation to update the current slowness model. From the perturbation theory, we have that $d = \widehat{d} + \Delta d$, $u = \widehat{u} + \Delta u$ and, consequently, $r = \widehat{r} + \Delta r$ are physical realizations with $s = \widehat{s} + \Delta s$, where the *hat* refers to fields obtained with the background slowness. To the first order (Born approximation), these perturbed fields are given by

$$\Delta d_{z+1}(\omega) = \mathbf{T}_z^{\downarrow}(\omega, \widehat{s})\Delta d_z(\omega) + \widetilde{\widetilde{\mathbf{D}}}_z(\omega)\Delta s_z \tag{13}$$

$$\Delta u_{z+1}(\omega) = \mathbf{T}_z^{\downarrow}(\omega, \widehat{s})\Delta u_z(\omega) + \widetilde{\widetilde{\mathbf{U}}}_z(\omega)\Delta s_z \tag{14}$$

The diagonal operators $\widetilde{\widetilde{\mathbf{D}}}_z$ and $\widetilde{\widetilde{\mathbf{U}}}_z$ have in the diagonal entries the scattered source and receiver wavefields, respectively, are given by the action of the scattering operator $\Delta\mathbf{T}_z^{\downarrow}$ on the background wavefields

$$\widetilde{\widetilde{\mathbf{D}}}_z(\omega) = \Delta\mathbf{T}_z^{\downarrow}(\omega, \widehat{s})\widehat{d}_z(\omega) = i\frac{\omega^2\widehat{s}}{\sqrt{\omega^2\widehat{s}^2 - |\mathbf{k}|^2}}dz\, \widehat{d}_z(\omega), \tag{15}$$

and

$$\widetilde{\widetilde{\mathbf{U}}}_z(\omega) = \Delta\mathbf{T}_z^{\downarrow}(\omega, \widehat{s})\widehat{u}_z(\omega) = -i\frac{\omega^2\widehat{s}}{\sqrt{\omega^2\widehat{s}^2 - |\mathbf{k}|^2}}dz\, \widehat{u}_z(\omega). \tag{16}$$