Appendix A

Supplemental Material for “Finite-frequency tomography using adjoint methods — Methodology and examples using membrane surface waves” (Chapter 2)

Note

Table A.1 makes a qualitative comparison between “classical” tomography and “adjoint” tomography. Table A.2 highlights all possible source-structure inversion experiments.

A.1 From misfit function to adjoint source: 2D membrane-wave example

Here we derive (2.48), following *Tromp et al.* (2005), which makes use of Green's functions. Alternatively, one could also use the Lagrange multiplier method (e.g., *Liu and Tromp*, 2006; *Fichtner et al.*, 2006).

For ease of notation, we let $\mathbf{x} = (x, y)$ and consider a single event with R recording receivers. The variation in the traveltime misfit function due to a model perturbation $\boldsymbol{\delta}\mathbf{m}$ is given by (2.7):

$$\delta F = -\sum_{r=1}^{R} \Delta T_r \, \delta T_r. \tag{A.1}$$

The cross-correlation traveltime variation δT_r can be written as (*Luo and Schuster*, 1990; *Marquering et al.*, 1999)

$$\delta T_r = \frac{1}{M_{\mathrm{T}r}} \int_0^T w_r(t) \partial_t s(\mathbf{x}_r, t) \delta s(\mathbf{x}_r, t) \, \mathrm{d}t, \tag{A.2}$$

where w_r denotes the cross-correlation window, δs the change in displacement, and $M_{\mathrm{T}r}$ the normalization factor defined as

$$M_{\mathrm{T}r} = \int_0^T w_r(t) s(\mathbf{x}_r, t) \partial_t^2 s(\mathbf{x}_r, t) \, \mathrm{d}t, \tag{A.3}$$

such that $M_{\mathrm{T}r} < 0$ for a pulse with nonzero amplitude.

The equation of motion that is solved by the SEM algorithm is shown in (2.29). Using the standard Green's function approach, we write the wavefield generated by the point source (2.30) as

$$s(\mathbf{x}, t) = \int_0^t \int_\Omega G(\mathbf{x}, \mathbf{x}'; t - t') \, f(\mathbf{x}', t') \, \mathrm{d}^2\mathbf{x}' \, \mathrm{d}t'. \tag{A.4}$$

The change in displacement δs due to a change in the point force δf may be written as

$$\delta s(\mathbf{x}, t) = \int_0^t \int_\Omega G(\mathbf{x}, \mathbf{x}'; t - t') \, \delta f(\mathbf{x}', t') \, \mathrm{d}^2\mathbf{x}' \, \mathrm{d}t'. \tag{A.5}$$

Upon substitution of the perturbation (A.5) into (A.1)–(A.2) we find that the change in the traveltime misfit function may be expressed as[1]

$$\begin{aligned}
\delta F &= \int_0^T \int_\Omega \delta f(\mathbf{x}, t') \left[\sum_{r=1}^{R} \Delta T_r \frac{1}{M_{\mathrm{T}r}} \int_0^{T-t'} G(\mathbf{x}, \mathbf{x}_r; T - t' - t)\, w_r(T-t) \partial_t s(\mathbf{x}_r, T-t)\, \mathrm{d}t \right] \mathrm{d}^2\mathbf{x}\, \mathrm{d}t' \\
&= \int_0^T \int_\Omega \delta f(\mathbf{x}, t)\, s^\dagger(\mathbf{x}, T-t)\, \mathrm{d}^2\mathbf{x}\, \mathrm{d}t, \qquad \text{(A.7)}
\end{aligned}$$

where we have defined the adjoint wavefield by

$$s^\dagger(\mathbf{x}, t) \equiv \int_0^t \int_\Omega G(\mathbf{x}, \mathbf{x}'; t - t')\, f^\dagger(\mathbf{x}', t')\, \mathrm{d}^2\mathbf{x}'\, \mathrm{d}t' \qquad \text{(A.8)}$$

and the adjoint source by

$$f^\dagger(\mathbf{x}, t) \equiv \sum_{r=1}^{R} \Delta T_r \frac{1}{M_{\mathrm{T}r}}\, w_r(T-t) \partial_t s(\mathbf{x}_r, T-t) \delta(\mathbf{x} - \mathbf{x}_r). \qquad \text{(A.9)}$$

Note that the spatial integration in (A.8) arises from the delta function in (A.9), and also that the adjoint source includes the time-reversed synthetic velocity recorded at the rth receiver.

A.2 The conjugate gradient algorithm

The gradient is not a vector but rather a tangent plane or set of level lines (*Tarantola*, 2005, p. 205). The metric (tensor) provides a means for selecting the steepest descent vector; using a different metric will lead to a different steepest descent vector. The metric also appears in the conjugate gradient algorithm, and thus the choices of metric will affect the optimization.

[1] The 3D version of Equation (A.7) is given by

$$\delta F = \int_0^T \int_V \delta \mathbf{f}(\mathbf{x}, t) \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)\, \mathrm{d}^3\mathbf{x}\, \mathrm{d}t\,. \qquad \text{(A.6)}$$

A.2.1 Background and notation

The model covariance matrix $\mathbf{C}$ defines the relationship between the gradient $\hat{\mathbf{g}}$ and the corresponding steepest ascent vector $\mathbf{g}$:

$$\mathbf{g} = \mathbf{C}\,\hat{\mathbf{g}}\,, \tag{A.10}$$

$$\hat{\mathbf{g}} = \mathbf{C}^{-1}\,\mathbf{g}. \tag{A.11}$$

Similarly, for the model vector,

$$\mathbf{m} = \mathbf{C}\,\hat{\mathbf{m}}\,, \tag{A.12}$$

$$\hat{\mathbf{m}} = \mathbf{C}^{-1}\,\mathbf{m}. \tag{A.13}$$

The L2-norm can be defined over either space:

$$\begin{aligned}\|\mathbf{m}\|_2 &= \left(\mathbf{m}^T\,\mathbf{C}^{-1}\,\mathbf{m}\right)^{1/2} = \left((\mathbf{C}\,\hat{\mathbf{m}})^T\;\mathbf{C}^{-1}\;(\mathbf{C}\,\hat{\mathbf{m}})\right)^{1/2} = \left(\hat{\mathbf{m}}^T\,\mathbf{C}^T\,\mathbf{C}^{-1}\,\mathbf{C}\,\hat{\mathbf{m}}\right)^{1/2} \\ &= \left(\hat{\mathbf{m}}^T\,\mathbf{C}\,\hat{\mathbf{m}}\right)^{1/2} = \left(\hat{\mathbf{m}}^T\,\hat{\mathbf{C}}^{-1}\,\hat{\mathbf{m}}\right)^{1/2} = \|\hat{\mathbf{m}}\|_2\,,\end{aligned} \tag{A.14}$$

where

$$\hat{\mathbf{C}} = \mathbf{C}^{-1}.$$

The duality product between the steepest ascent vector and the gradient can be written in several ways:

$$\langle \mathbf{g},\, \hat{\mathbf{g}}\rangle = \mathbf{g}^T\,\hat{\mathbf{g}} = \mathbf{g}^T\mathbf{C}^{-1}\,\mathbf{g} = \|\mathbf{g}\|_2^2\;, \tag{A.15}$$

$$\langle \mathbf{g},\, \hat{\mathbf{g}}\rangle = \mathbf{g}^T\hat{\mathbf{g}} = (\mathbf{C}\,\hat{\mathbf{g}})^T\,\hat{\mathbf{g}} = \hat{\mathbf{g}}^T\mathbf{C}\,\hat{\mathbf{g}} = \|\hat{\mathbf{g}}\|_2^2\;. \tag{A.16}$$

This shows that the norm of the steepest ascent vector is equal to the norm of the gradient, as expected.

A.2.2 Algorithm

The conjugate gradient algorithm we use may be summarized as follows: given an initial model $\mathbf{m}^0$, calculate $F(\mathbf{m}^0)$, $\hat{\mathbf{g}}^0 = \partial F/\partial \mathbf{m}(\mathbf{m}^0)$, and set the initial conjugate gradient search direction equal to minus the initial gradient of the misfit function,

$$\mathbf{p}^0 = -\mathbf{g}^0 = -\mathbf{C}\,\hat{\mathbf{g}}^0. \tag{A.17}$$

If $||\mathbf{p}^0|| < \epsilon$, where ϵ is a suitably small number, then $\mathbf{m}^0$ is the model we seek to determine, otherwise:

1. We denote a model in the direction of the search vector as, and its corresponding gradient, as

$$\mathbf{m}_\nu^k \equiv \mathbf{m}^k + \nu \mathbf{p}^k \,, \tag{A.18}$$

$$\hat{\mathbf{g}}_\nu^k \equiv \frac{\partial F}{\partial \mathbf{m}}\left(\mathbf{m}_\nu^k\right). \tag{A.19}$$

   Perform a line search to obtain the scalar ν^k that minimizes the function $\tilde{F}^k(\nu)$ where

$$\tilde{F}^k(\nu) = F(\mathbf{m}_\nu^k)\,, \tag{A.20}$$

$$\tilde{g}^k(\nu) = \frac{\partial \tilde{F}^k}{\partial \nu} = \left\langle \hat{\mathbf{g}}_\nu^k, \mathbf{p}^k \right\rangle = \left(\hat{\mathbf{g}}_\nu^k\right)^T \mathbf{p}^k \,. \tag{A.21}$$

   - Choose a test parameter $\nu_t^k = -2\tilde{F}^k(0)/\tilde{g}^k(0)$, based on quadratic extrapolation.
   - Calculate the test model $\mathbf{m}_t^k = \mathbf{m}^k + \nu_t^k \mathbf{p}^k$.
   - Calculate $F(\mathbf{m}_t^k)$ and, for cubic interpolation, $\hat{\mathbf{g}}_t^k = \partial F/\partial \mathbf{m}(\mathbf{m}_t^k)$.
   - Interpolate the function $\tilde{F}^k(\nu)$ by a quadratic or cubic polynomial and obtain the ν^k that gives the (analytical) minimum value of this polynomial.

2. Update the model: $\mathbf{m}^{k+1} = \mathbf{m}^k + \nu^k \mathbf{p}^k$, then calculate

$$\mathbf{g}^{k+1} = \mathbf{C}\,\hat{\mathbf{g}}^{k+1} = \mathbf{C}\,\frac{\partial F}{\partial \mathbf{m}}(\mathbf{m}^{k+1}). \tag{A.22}$$

3. Update the conjugate gradient search direction: $\mathbf{p}^{k+1} = -\mathbf{g}^{k+1} + \beta^{k+1}\mathbf{p}^k$, where

$$\beta^{k+1} = \frac{\left\langle \hat{\mathbf{g}}^{k+1} - \hat{\mathbf{g}}^k,\ \mathbf{g}^{k+1} \right\rangle}{\left\langle \hat{\mathbf{g}}^k, \mathbf{g}^k \right\rangle} = \frac{\left(\hat{\mathbf{g}}^{k+1} - \hat{\mathbf{g}}^k\right)^T \mathbf{C}\, \hat{\mathbf{g}}^{k+1}}{\left(\hat{\mathbf{g}}^k\right)^T \mathbf{C}\, \hat{\mathbf{g}}^k} \,. \tag{A.23}$$

4. If $||\mathbf{p}^{k+1}|| < \epsilon$, then $\mathbf{m}^{k+1}$ is the desired model; otherwise replace k with $k+1$ and restart from Step 1.

A.2.3 Inversion details of *Tape et al.* (2007)

Here we show how the description of the CG algorithm in *Tape et al.* (2007) leads to the general expressions in Section A.2. We use the tilde notation (e.g., $\widetilde{\mathbf{m}}$) to distinguish the notation in *Tape et al.* (2007) from the notation previously discussed.

From the CG algorithm (Section A.2.2), the first test model is given by

$$\widetilde{\mathbf{m}}_t^0 = \widetilde{\mathbf{m}}^0 + \widetilde{\nu}_t^0\, \widetilde{\mathbf{p}}^0 = \widetilde{\mathbf{m}}^0 - \widetilde{\nu}_t^0\, \widetilde{\mathbf{C}}\, \widetilde{\hat{\mathbf{g}}}^0 \,, \tag{A.24}$$

where the step length is

$$\widetilde{\nu}_t^0 = -\frac{2\,F\left(\widetilde{\mathbf{m}}^0\right)}{\left(\widetilde{\hat{\mathbf{g}}}^0\right)^T \widetilde{\mathbf{C}}\, \widetilde{\hat{\mathbf{g}}}^0} \,. \tag{A.25}$$

In *Tape et al.* (2007) we expanded the model into orthonormal basis functions and scaled the source parameters in a manner that allowed us to use

$$\widetilde{\mathbf{C}} = \mathbf{I} \tag{A.26}$$

in Equations (A.24) and (A.25).

The model vector and gradient vector are shown within the schematic expression for the (test) model update,

$$\widetilde{\boldsymbol{\delta}\mathbf{m}} \;=\; \widetilde{\mathbf{m}}^{0t} - \widetilde{\mathbf{m}}^0 = -\widetilde{\nu}_t^0\, \widetilde{\hat{\mathbf{g}}}^0 \,, \tag{A.27}$$

which is expanded as

$$\widetilde{\boldsymbol{\delta}\mathbf{m}} = \begin{bmatrix} C_1^{0t}\sqrt{V_1}/J \\ \vdots \\ C_i^{0t}\sqrt{V_i}/J \\ \vdots \\ C_H^{0t}\sqrt{V_H}/J \\ B_1^{0t}\sqrt{V_1}/J \\ \vdots \\ B_i^{0t}\sqrt{V_i}/J \\ \vdots \\ B_H^{0t}\sqrt{V_H}/J \\ (T_s)_1^{0t}\,/\,\widetilde{\sigma}_{t_s} \\ (X_s)_1^{0t}\,/\,\widetilde{\sigma}_{x_s} \\ (Y_s)_1^{0t}\,/\,\widetilde{\sigma}_{y_s} \\ (Z_s)_1^{0t}\,/\,\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_s^{0t}\,/\,\widetilde{\sigma}_{t_s} \\ (X_s)_s^{0t}\,/\,\widetilde{\sigma}_{x_s} \\ (Y_s)_s^{0t}\,/\,\widetilde{\sigma}_{y_s} \\ (Z_s)_s^{0t}\,/\,\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_S^{0t}\,/\,\widetilde{\sigma}_{t_s} \\ (X_s)_S^{0t}\,/\,\widetilde{\sigma}_{x_s} \\ (Y_s)_S^{0t}\,/\,\widetilde{\sigma}_{y_s} \\ (Z_s)_S^{0t}\,/\,\widetilde{\sigma}_{z_s} \end{bmatrix} - \begin{bmatrix} C_1^{0}\sqrt{V_1}/J \\ \vdots \\ C_i^{0}\sqrt{V_i}/J \\ \vdots \\ C_H^{0}\sqrt{V_H}/J \\ B_1^{0}\sqrt{V_1}/J \\ \vdots \\ B_i^{0}\sqrt{V_i}/J \\ \vdots \\ B_H^{0}\sqrt{V_H}/J \\ (T_s)_1^{0}\,/\,\widetilde{\sigma}_{t_s} \\ (X_s)_1^{0}\,/\,\widetilde{\sigma}_{x_s} \\ (Y_s)_1^{0}\,/\,\widetilde{\sigma}_{y_s} \\ (Z_s)_1^{0}\,/\,\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_s^{0}\,/\,\widetilde{\sigma}_{t_s} \\ (X_s)_s^{0}\,/\,\widetilde{\sigma}_{x_s} \\ (Y_s)_s^{0}\,/\,\widetilde{\sigma}_{y_s} \\ (Z_s)_s^{0}\,/\,\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_S^{0}\,/\,\widetilde{\sigma}_{t_s} \\ (X_s)_S^{0}\,/\,\widetilde{\sigma}_{x_s} \\ (Y_s)_S^{0}\,/\,\widetilde{\sigma}_{y_s} \\ (Z_s)_S^{0}\,/\,\widetilde{\sigma}_{z_s} \end{bmatrix} = -\widetilde{\nu}_t^0 \begin{bmatrix} K_1^{C}\sqrt{V_1}\,J \\ \vdots \\ K_i^{C}\sqrt{V_i}\,J \\ \vdots \\ K_H^{C}\sqrt{V_H}\,J \\ K_1^{B}\sqrt{V_1}\,J \\ \vdots \\ K_i^{B}\sqrt{V_i}\,J \\ \vdots \\ K_H^{B}\sqrt{V_H}\,J \\ K_1^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_1^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_1^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_1^{z_s}\,\widetilde{\sigma}_{z_s} \\ \vdots \\ K_s^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_s^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_s^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_s^{z_s}\,\widetilde{\sigma}_{z_s} \\ \vdots \\ K_S^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_S^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_S^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_S^{z_s}\,\widetilde{\sigma}_{z_s} \end{bmatrix}, \tag{A.28}$$

where J is a constant, and the values from *Tape et al.* (2007) are

$$\widetilde{\sigma}_{t_s} \equiv \tau = 20 \text{ s} , \tag{A.29}$$

$$\widetilde{\sigma}_{x_s} \equiv \lambda = 70,000 \text{ m} , \tag{A.30}$$

$$\widetilde{\sigma}_{y_s} \equiv \lambda = 70,000 \text{ m} , \tag{A.31}$$

$$\widetilde{\sigma}_{z_s} \equiv \lambda = 70,000 \text{ m}. \tag{A.32}$$

These terms are analogous to the uncertainties in the prior model parameters. For southern California tomography, reasonable values are

$$\sigma_{t_s} = 0.5 \text{ s} , \tag{A.33}$$

$$\sigma_{x_s} = 2000.0 \text{ m} , \tag{A.34}$$

$$\sigma_{y_s} = 2000.0 \text{ m} , \tag{A.35}$$

$$\sigma_{z_s} = 2000.0 \text{ m} . \tag{A.36}$$

We now define the scaling vector $\mathbf{w}$ as

$$\mathbf{w} \equiv \begin{bmatrix} J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \end{bmatrix}. \tag{A.37}$$

With $\mathbf{W} = \mathrm{diag}(\mathbf{w})$, we multiply Equation (A.28) by $\mathbf{W}$, and the (test) model update is then

$$\mathbf{W}\widetilde{\boldsymbol{\delta}\mathbf{m}} = \mathbf{W}\widetilde{\mathbf{m}}^{0t} - \mathbf{W}\widetilde{\mathbf{m}}^{0} = -\mathbf{W}\,\widetilde{\nu}_t^0\,\widetilde{\hat{\mathbf{g}}}^0\,,$$

$$\mathbf{W}\widetilde{\boldsymbol{\delta}\mathbf{m}} = \begin{bmatrix} C_1^{0t} \\ \vdots \\ C_i^{0t} \\ \vdots \\ C_H^{0t} \\ B_1^{0t} \\ \vdots \\ B_i^{0t} \\ \vdots \\ B_H^{0t} \\ (T_s)_1^{0t} \\ (X_s)_1^{0t} \\ (Y_s)_1^{0t} \\ (Z_s)_1^{0t} \\ \vdots \\ (T_s)_s^{0t} \\ (X_s)_s^{0t} \\ (Y_s)_s^{0t} \\ (Z_s)_s^{0t} \\ \vdots \\ (T_s)_S^{0t} \\ (X_s)_S^{0t} \\ (Y_s)_S^{0t} \\ (Z_s)_S^{0t} \end{bmatrix} - \begin{bmatrix} C_1^{0} \\ \vdots \\ C_i^{0} \\ \vdots \\ C_H^{0} \\ B_1^{0} \\ \vdots \\ B_i^{0} \\ \vdots \\ B_H^{0} \\ (T_s)_1^{0} \\ (X_s)_1^{0} \\ (Y_s)_1^{0} \\ (Z_s)_1^{0} \\ \vdots \\ (T_s)_s^{0} \\ (X_s)_s^{0} \\ (Y_s)_s^{0} \\ (Z_s)_s^{0} \\ \vdots \\ (T_s)_S^{0} \\ (X_s)_S^{0} \\ (Y_s)_S^{0} \\ (Z_s)_S^{0} \end{bmatrix} = -\widetilde{\nu}_t^0 \begin{bmatrix} J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \end{bmatrix} \begin{bmatrix} K_1^C\,\sqrt{V_1}\,J \\ \vdots \\ K_i^C\,\sqrt{V_i}\,J \\ \vdots \\ K_H^C\,\sqrt{V_H}\,J \\ K_1^B\,\sqrt{V_1}\,J \\ \vdots \\ K_i^B\,\sqrt{V_i}\,J \\ \vdots \\ K_H^B\,\sqrt{V_H}\,J \\ K_1^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_1^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_1^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_1^{z_s}\,\widetilde{\sigma}_{z_s} \\ \vdots \\ K_s^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_s^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_s^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_s^{z_s}\,\widetilde{\sigma}_{z_s} \\ \vdots \\ K_S^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_S^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_S^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_S^{z_s}\,\widetilde{\sigma}_{z_s} \end{bmatrix} = -\widetilde{\nu}_t^0 \begin{bmatrix} J^2\,K_1^C \\ \vdots \\ J^2\,K_i^C \\ \vdots \\ J^2\,K_H^C \\ J^2\,K_1^C \\ \vdots \\ J^2\,K_i^C \\ \vdots \\ J^2\,K_H^C \\ (\widetilde{\sigma}_{t_s})^2\,K_1^{t_s} \\ (\widetilde{\sigma}_{x_s})^2\,K_1^{x_s} \\ (\widetilde{\sigma}_{y_s})^2\,K_1^{y_s} \\ (\widetilde{\sigma}_{z_s})^2\,K_1^{z_s} \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2\,K_s^{t_s} \\ (\widetilde{\sigma}_{x_s})^2\,K_s^{x_s} \\ (\widetilde{\sigma}_{y_s})^2\,K_s^{y_s} \\ (\widetilde{\sigma}_{z_s})^2\,K_s^{z_s} \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2\,K_S^{t_s} \\ (\widetilde{\sigma}_{x_s})^2\,K_S^{x_s} \\ (\widetilde{\sigma}_{y_s})^2\,K_S^{y_s} \\ (\widetilde{\sigma}_{z_s})^2\,K_S^{z_s} \end{bmatrix}. \tag{A.38}$$

This can be rearranged as

$$\mathbf{W}\widetilde{\boldsymbol{\delta}\mathbf{m}} = \mathbf{W}\widetilde{\mathbf{m}}^{0t} - \mathbf{W}\widetilde{\mathbf{m}}^{0} = -\mathbf{W}\,\widetilde{\nu}_t^0\,\widetilde{\hat{\mathbf{g}}}^0 \,, \tag{A.39}$$

$$\boldsymbol{\delta}\mathbf{m} = \mathbf{m}^{0t} - \mathbf{m}^{0} = -\mathbf{W}\,\widetilde{\nu}_t^0\,\widetilde{\hat{\mathbf{g}}}^0 \tag{A.40}$$

$$\begin{aligned}\boldsymbol{\delta}\mathbf{m} &= -\widetilde{\nu}_t^0 \begin{bmatrix} J^2/V_1 \\ \vdots \\ J^2/V_i \\ \vdots \\ J^2/V_H \\ J^2/V_1 \\ \vdots \\ J^2/V_i \\ \vdots \\ J^2/V_H \\ (\widetilde{\sigma}_{t_s})^2 \\ (\widetilde{\sigma}_{x_s})^2 \\ (\widetilde{\sigma}_{y_s})^2 \\ (\widetilde{\sigma}_{z_s})^2 \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2 \\ (\widetilde{\sigma}_{x_s})^2 \\ (\widetilde{\sigma}_{y_s})^2 \\ (\widetilde{\sigma}_{z_s})^2 \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2 \\ (\widetilde{\sigma}_{x_s})^2 \\ (\widetilde{\sigma}_{y_s})^2 \\ (\widetilde{\sigma}_{z_s})^2 \end{bmatrix} \begin{bmatrix} K_1^C\,V_1 \\ \vdots \\ K_i^C\,V_i \\ \vdots \\ K_H^C\,V_H \\ K_1^B\,V_1 \\ \vdots \\ K_i^B\,V_i \\ \vdots \\ K_H^B\,V_H \\ K_1^{t_s} \\ K_1^{x_s} \\ K_1^{y_s} \\ K_1^{z_s} \\ \vdots \\ K_s^{t_s} \\ K_s^{x_s} \\ K_s^{y_s} \\ K_s^{z_s} \\ \vdots \\ K_S^{t_s} \\ K_S^{x_s} \\ K_S^{y_s} \\ K_S^{z_s} \end{bmatrix} \\ &= -\widetilde{\nu}_t^0\,\mathbf{C}'\,\hat{\mathbf{g}}^0 \,. \end{aligned} \tag{A.41}$$

Thus, the "old" method is equivalent to the new method, but using a diagonal covariance matrix defined by $\mathbf{C}'$ instead of by Equation (A.42). Furthermore, $\widetilde{\nu}_t^0$ will differ from ν_t^0, since a difference covariance matrix is present.

$$\mathbf{c} \;=\; \begin{bmatrix} \sigma_C^2\, V\,/\,V_1 \\ \vdots \\ \sigma_C^2\, V\,/\,V_i \\ \vdots \\ \sigma_C^2\, V\,/\,V_H \\ \sigma_B^2\, V\,/\,V_1 \\ \vdots \\ \sigma_B^2\, V\,/\,V_i \\ \vdots \\ \sigma_B^2\, V\,/\,V_H \\ \sigma_{t_s}^2 \\ \sigma_{x_s}^2 \\ \sigma_{y_s}^2 \\ \sigma_{z_s}^2 \\ \vdots \\ \sigma_{t_s}^2 \\ \sigma_{x_s}^2 \\ \sigma_{y_s}^2 \\ \sigma_{z_s}^2 \\ \vdots \\ \sigma_{t_s}^2 \\ \sigma_{x_s}^2 \\ \sigma_{y_s}^2 \\ \sigma_{z_s}^2 \end{bmatrix} . \qquad \text{(A.42)}$$

Table A.1: Comparison between two generic tomography approaches, 'classical' and 'adjoint'. The reference model is expanded in terms of basis functions $B_k(\mathbf{x})$. $K_i(\mathbf{x})$ denotes the data-independent kernel for the ith source-receiver pair, while $K(\mathbf{x})$ denotes the data-dependent misfit kernel computed via adjoint methods.

	classical tomography	adjoint tomography
reference model	1D	3D
physical domain	3D	3D
Born approximation	yes	yes
forward modelling technique	e.g., ray theory, modes, or banana-doughnut kernels	fully numerical (e.g., SEM)
gradient method	$\mathbf{g} = -\mathbf{G}^T\mathbf{d}$ $G_{ik} = \int_V K_i B_k \, \mathrm{d}^3\mathbf{x}$	$g_k = \int_V K B_k \, \mathrm{d}^3\mathbf{x}$
Newton method	$\mathbf{G}^T\mathbf{G}\,\delta\mathsf{m} \approx -\mathbf{g}$	(too costly)
number of iterations	1	multiple

Table A.2: Source inversions, structure inversions, and joint inversions. T07 = *Tape et al.* (2007).

	type	pert source	pert structure	invert source	invert structure	comments	T07 figure
1	source	Y	N	Y	N	basic source inversion	16
2	structure	N	Y	N	Y	basic structure inversion	17a–c
3	joint	Y	Y	Y	Y	basic joint inversion	17d–f
4	structure	Y	Y	N	Y	map src error to structure	19a–c
5	structure	Y	N	N	Y	map src error to structure	(none)
6	source	Y	Y	Y	N	map structure error to src	19d–f
7	source	N	Y	Y	N	map structure error to src	(none)
8	joint	Y	N	Y	Y	map src error to structure	(none)
9	joint	N	Y	Y	Y	map structure error to src	(none)