Comparison of Several Ways to Compute Sensitivity Kernels

Bill Menke, December 16, 2022

1. Introduction

Consider a linear differential equation of the form, \( \mathcal{L}(c) u = s \), where the field, \( u(x, y, z, t) \), and the source, \( s(x, y, z, t) \), depend on position, \( (x, y, z) \), and time, \( t \), and where the differential operator, \( \mathcal{L}(c) \), depends on a scalar material parameter, \( c \). An example is a wave propagation problem, involving pressure, \( u \), source, \( s \), background velocity, \( v \), and heterogeneity, \( cf(x, y, z) \)

\[
\mathcal{L} = (v^2 + cf(x, y, z)) \frac{\partial^2}{\partial t^2} - \nabla^2
\]

Given time-dependent measurements, \( u_i^{obs}(t) = u^{obs}(x_i, y_i, z_i, t) \) at \( N \) observer positions (stations), \( (x_i, y_i, z_i) \), the individual errors are defined as \( e_i = u_i^{obs} - u_i \), and the total error as

\[
E = \int \sum_i e_i^2 \, dt
\]

The term sensitivity kernel means either the partial derivative, \( \partial e_i / \partial c \) (the sensitivity of individual errors to perturbations in the material parameter) or \( \partial E / \partial c \) (the sensitivity kernel of the total error to perturbations in the material parameter), depending on context. Sensitivity kernels can be computed in a variety of ways.

Both sensitivity kernels depend on the quantity, \( \partial u_i / \partial c \)

\[
\frac{\partial e_i}{\partial c} = \frac{\partial}{\partial c} (u_i^{obs} - u_i) = -\frac{\partial u_i}{\partial c}
\]

\[
\frac{\partial E}{\partial c} = \frac{\partial}{\partial c} \int \sum_i e_i^2 \, dt = -\int \sum_i 2e_i \frac{\partial u_i}{\partial c} \, dt
\]

The following mathematical trick helps to “hide” the stations in the formula for \( \partial E / \partial c \). We imagine that the error, \( e = u_i^{obs} - u \) is known everywhere. Then,

\[
\frac{\partial E}{\partial c} = -\int \sum_i 2e_i \frac{\partial u_i}{\partial c} \, dt = -\left(2q, \frac{\partial u}{\partial c}\right) \text{ with } q = \sum_i e_i \delta(x - x_i)\delta(y - y_i)\delta(z - z_i)
\]

Here \((.,.\)\) is the inner product, defined as

\[
(a, b) \equiv \int \int \int a \, b \, d^3x \, dt
\]

2. Methods of computing \( \partial u / \partial c \)

2.1. Method A. Finite differences
Suppose that the differential equation, \( \mathcal{L}(c)u = s \) is solved twice, once for material parameter, \( c_0 \), leading to solution, \( u(c_0) \) and another for \( c_0 + \Delta c \), leading to solution, \( u(c_0 + \Delta c) \). The derivative, \( \partial u / \partial c \), can be approximated as

\[
\frac{\partial u_0}{\partial c} \equiv \left. \frac{\partial u}{\partial c} \right|_{c_0} \approx \frac{u(c_0 + \Delta c) - u(c_0)}{\Delta c}
\]

And then \( \partial u_i / \partial c \) is just \( \partial u_0 / \partial c \) evaluated at \((x_i, y_i, z_i)\).

Superficially, this finite difference derivative hides the “interaction” that makes \( u(c_0 + \Delta c) \) different from \( u(c_0) \). In a seismic problem, the interaction corresponds to the scattering of the wavefield off of the heterogeneity. However, the interaction can be teased out of this formula by substituting \( u = \mathcal{L}^{-1}s \)

\[
\frac{\partial u_0}{\partial c} \approx \frac{[\mathcal{L}(c_0 + \Delta c)]^{-1} - [\mathcal{L}(c_0)]^{-1}}{\Delta c} = \left. \frac{\partial \mathcal{L}^{-1}}{\partial c} \right|_{c_0} s = -\mathcal{L}_0^{-1} \frac{\partial \mathcal{L}_0}{\partial c} \mathcal{L}_0^{-1} s = -\mathcal{L}_0^{-1} \frac{\partial \mathcal{L}_0}{\partial c} u_0
\]

Here, we have defined \( u_0 \equiv u(c_0) \), \( \mathcal{L}_0 \equiv \mathcal{L}(c_0) \) and \( \partial \mathcal{L}_0 / \partial c \equiv \partial \mathcal{L} / \partial c |_{c_0} \) and have used the rule,

\[
\frac{\partial \mathcal{L}^{-1}}{\partial c} = -\mathcal{L}^{-1} \frac{\partial \mathcal{L}}{\partial c} \mathcal{L}^{-1}
\]

(which is derived by differentiating \( \mathcal{L} \mathcal{L}^{-1} = \mathcal{I} \), where \( \mathcal{I} \) is the identity operator). Thus, \( \partial u_0 / \partial c \) satisfies a differential equation

\[
\frac{\partial \mathcal{L}_0}{\partial c} u_0 \approx -\frac{\partial \mathcal{L}_0}{\partial c} u_0
\]

the r.h.s. of which is the “virtual source” or the “scattering interaction”.

2.2 Method B. Born approximation

We consider unperturbed and perturbed differential equations:

\[
\mathcal{L}(c_0) u(c_0) = s \\
\mathcal{L}(c_0 + \Delta c) u(c_0 + \Delta c) = s
\]

We expand both \( \mathcal{L}(c_0 + \Delta c) \) and \( u(c_0 + \Delta c) \) in Taylor series

\[
\left( \mathcal{L}_0 + \frac{\partial \mathcal{L}_0}{\partial c} \Delta c \right) \left( u_0 + \frac{\partial u}{\partial c} \Delta c \right) = s
\]

Multiply out and discard the higher order term

\[
\mathcal{L}_0 u_0 + \mathcal{L}_0 \frac{\partial u}{\partial c} \Delta c + \frac{\partial \mathcal{L}_0}{\partial c} \Delta c u_0 = s
\]

Subtract the unperturbed equation
\[
\mathcal{L}_0 \frac{\partial u}{\partial c} \Delta c + \frac{\partial \mathcal{L}_0}{\partial c} \Delta u_0 = 0
\]

And rearrange, yielding the same formula as before.

\[
\frac{\partial u_0}{\partial c} \equiv \frac{\partial u}{\partial c} \bigg|_{c_0} = -\mathcal{L}_0^{-1} \frac{\partial \mathcal{L}_0}{\partial c} u_0
\]

2.3. Method C, direct differentiation

We differentiate

\[ u(c) = \mathcal{L}^{-1}(c)s \]

to get

\[ \frac{\partial u}{\partial c} = \frac{\partial \mathcal{L}^{-1}}{\partial c} s = -\mathcal{L}^{-1} \frac{\partial \mathcal{L}_0}{\partial c} \mathcal{L}^{-1} s = -\mathcal{L}^{-1} \frac{\partial \mathcal{L}}{\partial c} u \]

or

\[ \frac{\partial u_0}{\partial c} = -\mathcal{L}_0^{-1} \frac{\partial \mathcal{L}_0}{\partial c} u_0 \]

Reassuringly, all three methods give identical results.

3. Methods of computing \( \partial E / \partial c \)

3.1 Direct methods

Step A. Compute \( u_0 \) everywhere by solving \( \mathcal{L}_0 u_0 = s \), and sample it to get its values at the receivers.

Step B. Compute \( \partial u_0 / \partial c \) everywhere, either by finite differences or by solving

\[ \mathcal{L}_0 \frac{\partial u_0}{\partial c} \approx -\frac{\partial \mathcal{L}_0}{\partial c} u_0 \]

and sample it to get its values at the receivers.

Step C. Perform the integrals/summation

\[ \frac{\partial E}{\partial c} = -\int \sum_i 2e_i \frac{\partial u_i}{\partial c} \, dt \]

Irrespective of details, two differential equations must be solved and one integration/summation must be performed to compute \( \partial E / \partial c \). In practice, the heterogeneities are described by many (say \( M \)) \( c \)s, so that \( 2M \) solutions and \( 1M \) integrations/summations must be performed.
3.2 Adjoint method

Step A. Substitute

\[ \frac{\partial u_0}{\partial c} = -\mathcal{L}_0^{-1} \frac{\partial \mathcal{L}_0}{\partial c} u_0 \]

into

\[ \frac{\partial E}{\partial c} = -\left( 2q, \frac{\partial u_0}{\partial c} \right) \]

to get

\[ \frac{\partial E_0}{\partial c} = \left( 2q, \mathcal{L}_0^{-1} \frac{\partial \mathcal{L}_0}{\partial c} u_0 \right) \]

Step B. Use the adjoint method to move \( \mathcal{L}_0^{-1} \) to other side of inner product

\[ \frac{\partial E_0}{\partial c} = \left( 2\mathcal{L}_0^+ q, \frac{\partial \mathcal{L}_0}{\partial c} u_0 \right) \]

Here, we have used fact that adjoint of an inverse is the inverse of an adjoint.

Step C. Now define, \( \lambda \equiv \mathcal{L}_0^+ q \) so that the “adjoint field” \( \lambda \) satisfies the adjoint differential equation

\[ \mathcal{L}_0^+ \lambda \equiv q \]

and solve this equation for \( \lambda \).

Step D. Perform the inner product

\[ \frac{\partial E_0}{\partial c} = \left( 2\lambda, \frac{\partial \mathcal{L}_0}{\partial c} u_0 \right) \]

This looks complicated, so why do people do it? The answer is all the information about the heterogeneity is in the \( \partial \mathcal{L}_0 / \partial c \) factor. Irrespective of the number of heterogeneities, you need perform only two solutions of differential equations, one for \( u_0 \) and one for \( \lambda \). You still need to perform \( M \) inner products, one for each distinct \( \partial \mathcal{L}_0 / \partial c \), but that’s the easy part.

In summary, the adjoint method provides a method for computing \( \partial E_0 / \partial c \) than is much more computationally efficient than direct methods.