For those who doubt the adjoint method can be applied to glacial isostatic problems
Bill Menke, November 12, 2023

Consider the assertion, "The quadratic formula fails when applied to seismological problems". By quadratic formula, I mean the one that one learns in high school:

$$\text{if} \quad ax^2 + bx + c = 0 \quad \text{then} \quad x = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$$

This result can be proved by inserting the formula for x into the quadratic equation and simplifying the terms.

$$\frac{ab^2}{4a^2} \mp \frac{2ab\sqrt{b^2 - 4ac}}{4a^2} + a\frac{(b^2 - 4ac)}{4a^2} - \frac{b^2}{2a} \pm \frac{b\sqrt{b^2 - 4ac}}{2a} + c \stackrel{?}{=} 0$$

$$\frac{b^2}{4a} \mp \frac{b\sqrt{b^2 - 4ac}}{2a} + \frac{b^2}{4a} - c - \frac{2b^2}{4a} \pm \frac{b\sqrt{b^2 - 4ac}}{2a} + c \stackrel{?}{=} 0$$

$$0 = 0$$

The result is completely general for the domain of complex numbers, without any qualification concerning the meaning of the unknown, x. So the assertion that it fails for one application (say, seismology) and not for another (say, high school algebra) is surprising; one should not accept it without a very careful argument attesting to its veracity. Of course, this is just a hypothetical; in fact, the quadratic formula does *not* fail when applied to seismology.

The quadratic formula, above, does not necessarily apply to other domains. It will fail for the domain of, e.g., matrices, because the above proof requires that some of the products, e.g., b and $\sqrt{b^2 - 4ac}$, commute, whereas matrix multiplication is non-commutative. Parenthetically, I note that the extension to the matrix quadratic equation is known, bears some similarity to the quadratic formula, but is far more complicated (Yuan et al., 2021, Applied Mathematics and Computation 410, 126463).

The situation is the same for the assertion, "The adjoint method fails to compute correct sensitivity kernels when applied to glacial isostatic problems". It should be viewed very skeptically in the absence of a very convincing rationale. The adjoint method to which I am referring is:

$$\text{if} \quad E = (\mathbf{h}, \mathbf{u}) \quad \text{and} \quad \boldsymbol{\mathcal{L}}(\mathbf{m})\mathbf{u} = \mathbf{f}$$

$$\text{then} \quad \left.\frac{\partial E}{\partial m_i}\right|_{\mathbf{m}_0} = \left(\boldsymbol{\lambda}, \boldsymbol{\xi}^{(i)}\right) \quad \text{with} \quad \boldsymbol{\mathcal{L}}(\mathbf{m}_0)^\dagger \boldsymbol{\lambda} = \mathbf{h} \quad \text{and} \quad \boldsymbol{\xi}^{(i)} = -\left.\frac{\partial \boldsymbol{\mathcal{L}}}{\partial m_i}\right|_{\mathbf{m}_0} \mathbf{u}(\mathbf{m}_0)$$

Here, $\boldsymbol{\mathcal{L}}(\mathbf{m})\mathbf{u} = \mathbf{f}$ is a linear partial differential equation for a vector field, $\mathbf{u}(x, y, z, t, \mathbf{m})$, considered to depend upon parameters, $\mathbf{m}$, and E is a scalar quantity that depends on $\mathbf{u}$ via an inner product with a known function, $\mathbf{h}$. The sensitivity kernel, $\left.\frac{\partial E}{\partial m_i}\right|_{\mathbf{m}_0}$, allows one to calculate the change in E caused by a perturbation, Δm_i, of one of the parameters away from the reference value, $\mathbf{m}_0$.

As with the quadratic formula, the proof does not depend upon the interpretation of $\mathbf{u}$, $\mathbf{h}$, or E, or on the dimension of $\mathbf{u}$. Here it is:

$$\frac{\partial E}{\partial m_i} = \left(\mathbf{h}, \frac{\partial \mathbf{u}}{\partial m_i}\right) = \left(\mathbf{h}, \frac{\partial}{\partial m_i}(\boldsymbol{\mathcal{L}}^{-1}\mathbf{f})\right) = \left(\mathbf{h}, \frac{\partial \boldsymbol{\mathcal{L}}^{-1}}{\partial m_i}\mathbf{f}\right)$$

$$= \left(\mathbf{h}, -\boldsymbol{\mathcal{L}}^{-1}\frac{\partial \boldsymbol{\mathcal{L}}}{\partial m_i}\boldsymbol{\mathcal{L}}^{-1}\mathbf{f}\right) = \left(\boldsymbol{\mathcal{L}}^{-1\dagger}\mathbf{h}, -\frac{\partial \boldsymbol{\mathcal{L}}}{\partial m_i}\mathbf{u}\right) = \left(\boldsymbol{\mathcal{L}}^{\dagger-1}\mathbf{h}, -\frac{\partial \boldsymbol{\mathcal{L}}}{\partial m_i}\mathbf{u}\right) = \left(\boldsymbol{\lambda}, \boldsymbol{\xi}^{(i)}\right)$$

Here, I have used the well-known formula for the derivative of an inverse operator, which can be obtained by using the chain rule to differentiate the equation, $\boldsymbol{\mathcal{L}}^{-1}\boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{I}}$, where $\boldsymbol{\mathcal{I}}$ is the identity operator. The only assumption made is that the operator, $\boldsymbol{\mathcal{L}}$, is invertible; that is, that the partial differential equation has a unique solution (a condition which is true for the glacial isostatic problem). Parenthetically, I mention that when the equation is not uniquely invertible, the only change in the method is the need for the additional constraint that $\mathbf{u}$ minimizes E, which leads to the a different equation for $\boldsymbol{\lambda}$, namely, $\boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{L}}^{\dagger}\boldsymbol{\lambda} = \boldsymbol{\mathcal{L}}\mathbf{h}$. The divergence is one such operator. The equation, $\nabla \cdot \mathbf{u} = 0$ (where $\mathbf{u}$ now is a 3-vector), is un-invertible because it does not constrain the curl of $\mathbf{u}$.

Given that one can prove the general case, I believe that the assertion the adjoint method doesn't work in glacial isostatic problems to be false. However, it is worth considering why it might *seem* to be true.

One possibility is that a sensitivity kernel calculated by the adjoint method seems different from one calculated by some other means. In such a case, I would be extremely attentive to whether the two "sensitivity kernels" actually refer to the same derivative. Consider that although $\partial A/\partial m_i$ and $\partial B/\partial m_i'$ are both sensitivity kernels, they are not equal unless $B = A$ and $m_i' = m_i$. Most geophysical problems are sufficiently complicated that even variables with the same letter names, but in different publications, may refer to different quantities (though, more often than not, closely related ones). Sensitivity kernels obey the chain rules:

$$\frac{\partial B}{\partial m_i} = \frac{\partial \mathrm{A}}{\partial m_i}\frac{\partial B}{\partial A} \quad \text{and} \quad \frac{\partial A}{\partial m_i'} = \sum_j \frac{\partial \mathrm{A}}{\partial m_j}\frac{\partial m_j}{\partial m_i'}$$

so it is possible (at least in principle) to convert one to another to facilitate an inter-comparison.

A related possibility concerns the *positions* at which the sensitivity kernels are evaluated. A parameter, say, m_j might be associated with a position, say (x_j, y_j, z_j). And the quantities, E and $\mathbf{h}$, might be associated with a different position, say (x_i, y_i, z_i) (in which case we should index them as $E^{(i)}$ and $\mathbf{h}^{(i)}$). A plot of the sensitivity kernel as a function of variable (x_j, y_j, z_j) and fixed (x_i, y_i, z_i) is *not* the same as a plot with variable (x_i, y_i, z_i) and fixed (x_j, y_j, z_j).