**Review of “Seismic image-domain wave-equation-based methods – an overview”, for *Geophysics***

This paper has the ambitious goal implicit in the title. I believe that it mostly succeeds. The text is somewhat breathless, probably unavoidable for traversing such a large tract of material. Yet the major ideas and contributions, and their relationships, are for the most part clearly described, with good illustrations. The careful reader will want to dip frequently into the cited literature, and therein lies much of the value of this manuscript, as an annotated bibliography.

I note below a few issues that I feel need to be addressed before acceptance for publication.

———————————-

P. 4 l. 30: “Gockenbach” is mis-spelled in the reference Gockenbach and Symes 1995.

P. 4 l. 44: “…they have the capabilities to handle finite frequency phenomena.” This point is made in several places, but I don’t find an explanation. What does “handle” mean? What finite-frequency phenomena? Do not wave equation methods have other advantages, for example geometric flexibility and/or computational efficiency, that have nothing directly to do with frequency content?

P. 5 l. 57: the adjoint of the Born (linearized) modeling operator is a reverse-time migration operator, but other RTM operators (not the Born adjoint) have been used, as have other migration operators (not RTM), for construction of MVA algorithms. The Born adjoint is only one possibility, and probably not even the most popular (I am not sure how one would take a vote!). This fact is explained later (p. 8) but the introduction of the Born adjoint here is confusing.

P. 8 l. 52: Kern and Symes 1994 is a better reference here than Shen and Symes 2008 (who used subsurface offset and one-way migration).

p. 9 l. 20: it might be better to explain “handle” in this context: that is, can produce a faithful image of reflectivity, without the phantom reflectors produced by surface-extended migration (which may not be focused (flat) even with synthetic data and correct velocity).

p. 10 eqn.8: the notation suggests that x and h have the same dimensionality (both bold-faced), whereas in the most widely used variant (first suggested by Claerbout) h is confined to a horizontal plane (one dimension less). How to constrain the subsurface offset to give a model with the same dimension as the data, while also allowing for a full spherical scattering aperture, is a subject in itself, and an ongoing research topic. This issue should be clearly described.

pp. 14-15: This effort to explain how the MVA gradient generates a long wavelength velocity update from short wavelength data is welcome, but not convincing. There is almost no distinction between the MVA gradient description on p. 14 (Figure 5) and the similar description of the FWI gradient (Figure 6). Note that Figures 5(a) and 6(b) are exactly the same - the formation of the reflector image happens in exactly the same way in both. Moreover if one imagines Figures 5(b) and 5(c) superimposed, one obtains Figure 6(c). It would appear just from this discussion and the accompanying Figures that, apart from diving wave tomography, there is little difference between MVA and FWI gradients, which does not reassure one that the MVA might be more convex, or less prone to cycle-skipping, than FWI.

In fact two elements of the MVA definition are completely missing from this schematic discussion: the extended model and the annihilator. The role of the extended model is simple to explain: as mentioned earlier, it makes the data-to-reflectivity relation invertible, so that data can always be fit, even with the wrong velocity model. For this reason, the extended reflectivity constructed by migrating or inverting the data tends to have a lot of energy even for poor velocity, whereas without the extension, destructive interference occurs for poor velocities and the reflectivity has little energy. This fact accounts for the weakness of the FWI gradient far from convergence. The role of the extension in removing this obstacle is very important, and does not show up in the discussion on pp. 14-15. Also, the annihilator seems to play no role, which cannot be correct.

I think that the authors can do much better than this at the very important job of conveying WHY the MVA (or IVA) approach can work. I urge them to try again, perhaps using a simple example that can be dissected “on paper”.

p. 14 l. 52: evidently a reference to Figure 6(b) belongs in here somewhere.

P. 16 l. 52 and numerous other places: the term “pseudo inverse” already has a precise and widely accepted definition in Linear Algebra. The simplest example is the least squares inverse of a matrix with independent columns. The “asymptotic wave equation” approximate inverse construction referenced here is not a pseudo inverse in this sense. I strongly recommend using “approximate inverse” instead.

P. 17 paragraph beginning l. 27: actually the principal cause of MVA gradient error, for the version using “true amplitude” reflectivity via iterative approximation of the reflectivity, is NOT the amplification of large offset extended reflectivity error by the annihilator. In fact, if that were the main cause, then you would expect the same type of error for the non-iterative “approximate inverse” approach explained in the next section, since it also involves an approximation to the “true” reflectivity, especially at large subsurface offsets. In fact, the same instability can be seen in the FWI gradient computed via the Krylov-Gauss-Newton approach (if one looks). The reason is that all of these calculations are instances of the Variable Projection Method, and of the use of the formula for the VPM gradient that assumes exact solution of the normal equation (least squares solution for the reflectivity). If the reflectivity is in error at all, then a second term appears in the VPM gradient, which involves the second derivative of the modeling operator. This second derivative includes an extra factor of frequency (or time/space derivatives) beyond the first derivative appearing in the gradient formula, and the least squares residual (or reflectivity gradient). For iterative inversion of the reflectivity, the latter tends to zero in the mean square sense, but that does not imply that its derivatives converge. Therefore the macro-model gradient does not necessarily converge: this is the instability mentioned in the text. For asymptotic approximate inversion of the reflectivity, on the other hand, apart from aperture effects and the like, the reflectivity error has lower order in frequency, which compensates for the extra factors of frequency in the modeling operator derivative. This effect accounts for the more stable behavior of the gradient based on approximate inverse operators, reported by the authors.

P. 40: the review of open issues is excellent, but leaves out one very interesting question: how can anelastic attenuation be accommodated? Attenuation is pervasive and non-negligible, and since it is accompanied by dispersion, can interact with velocity estimation. Note that all of the justifications for MVA are ultimately tomographic, that is, rely on ray theory, and that the “direct” inverse that plays so big a role in the MVA state-of-the-art is an asymptotic inverse, computed by wave equation calculations. However in the presence of positive Q, there is no energy in the wavefield, in the infinite-frequency limit. What can substitute for ray theory in explaining the kinematics of wavefields, on which MVA ultimately rests, and at the same time accommodate significant attenuation, as occurs in the sedimentary section?