

To the Editor,

Re: *Automatic adjoints without the algebra: flexible sensitivity analysis of nonlinear problems in the geosciences* by Jacob Frasunkiewicz, Christian Schuler, Evangelos Moulas and Boris J. P. Kaus, submitted to *Geoscientific Model Development* (egusphere-2026-3963).

Thank you for the opportunity to review this manuscript. The authors present a Julia framework in which forward-mode automatic differentiation is used to assemble the Jacobian of a discretised residual and its derivative with respect to material parameters. The transposed Jacobian is then used to solve an adjoint system and calculate sensitivity kernels. The method is demonstrated on steady diffusion, transient heat conduction, coupled flow and heat transport, a rising mantle plume and visco-elasto-plastic shear localisation. The manuscript also offers a geometric interpretation of the discrete adjoint calculation.

The underlying idea is sound, and a compact collection of reproducible examples could make a useful contribution to *Geoscientific Model Development*. The discrete adjoint algebra in Section 3.2 is correct, the one-dimensional sensitivity agrees with its analytical reference, and the archived code is sufficiently complete to inspect and run. However, the principal claims of the manuscript are not supported by what is computed. The transient kernels are not gradients of the stated time-dependent problems, one of the principal examples does not solve the adjoint equation, verification is too limited to establish accuracy, and no evidence is given for the claims of computational efficiency or scalability. There are also repeated disagreements between the equations, code, figures, captions and physical interpretations.

Several examples and their interpretation would need to be rebuilt, after which the paper's central claims and exposition would also need to change. The sections below set out what I think that would involve.

The shown “transient” kernels are not gradients of the posed problems

Sections 2 and 3 derive the adjoint for a stationary algebraic system, $R(u, m) = 0$. Four of the applications are time-dependent problem, where the problem is advanced in time with an implicit scheme. At time step n , the residual depends on both the current state and the state at step $n - 1$. The latter also depends on the parameters. Consequently, the derivative of an objective evaluated after many time steps requires an adjoint recursion through the *entire trajectory*.

The implementation I see in the scripts, is not actually doing this. It evaluates J and B for the current solve while *holding the previous state fixed*, and then solves one linear system. The previous-step fields are stored as ordinary floating-point arrays. The kernels in Figures 5, 6, 8 and 11 are therefore derivatives of one “solve block” with a frozen history. They are not derivatives of the stated objectives with respect to the parameters of the time-dependent model.

This distinction seems to be briefly acknowledged in the Discussion, where the kernels are called diagnostic, but it conflicts with the main claim of the paper, in the abstract, Equation 13, the application sections and the conclusions. In particular, the geothermal kernel cannot predict the effect of permeability on power after one hundred years when the dependence of the preceding one hundred years on permeability is excluded. Likewise, the plume kernel is not an inversion direction for a time-dependent plume trajectory.

The verification in Appendix A is not necessarily showing if the authors have done a proper calculation of the adjoint. The finite-difference calculation in Appendix A also starts from the stored previous state and re-solves only the current step. Agreement between that calculation and the adjoint verifies the differentiation and linear algebra of one frozen step. It does not verify the gradient of the transient problem. The authors must either implement and test the backward-in-time adjoint or state consistently, beginning in the abstract, that these are single-step partial derivatives and restrict their interpretation accordingly.

The numerical verification is not nearly enough

A model paper of this kind needs to establish three things: that the forward solver is accurate, that the adjoint equation is implemented correctly, and that the resulting derivative is consistent with the discrete forward model. The manuscript provides each of these only in a limited form. The one-dimensional example checks a steady diffusion calculation. The Appendix tests one frozen step of the shear-band model at one resolution and along one perturbation direction. The capsule, geothermal and plume examples have no derivative verification, and none of the transient examples has a grid or time-step convergence study for either the forward solution or the kernel.

The reporting of the Appendix also overstates how robust the verifications are. Figure A1 contains differences of approximately 10^{-6} to 10^{-5} for kernels with amplitudes of order 10^{-2} to 10^{-1} , which does not support the claim of seven to eight significant digits in the worst cells. Figure A2 cannot resolve errors at the claimed level. The floor in Figure A3 is attributed to round-off without reporting the objective, the projected gradient or the

nonlinear-solver error needed to assess that explanation. The manuscript gives a Newton tolerance of 10^{-8} , whereas the archived scripts use different tolerances for the production, finite-difference and gradient-test runs.

The test also avoids the non-differentiable switches that are the trickiest bits in automatic differentiation. A relative perturbation of approximately 1.5×10^{-8} is unlikely to move cells across the plastic yield surface. Similar qualifications apply to upwind selection, the tensile cap, the phase threshold used by the plume objective and changes in the adaptive time step. Automatic differentiation returns the derivative of the branch that is executed; it does not establish differentiability across a change of branch. This limitation is manageable, but it must be stated and tested.

There is a still more direct error in the plume example. The plume copy of the adjoint routine solves $J\psi = \partial F / \partial u$, not $J^T \psi = \partial F / \partial u$. The unknowns are coupled velocity, pressure and temperature fields, and, unless I am mistaken, this Jacobian is not symmetric. The fields in Figure 8 are therefore not adjoint sensitivities. The finite-difference path that would expose this error is disabled in the plume driver. Figure 8 and its physical interpretation cannot be assessed until the transpose is restored and the corrected result is independently verified.

The contribution, computational cost and range of application remain unclear

Differentiating a residual with automatic differentiation is no longer, by itself, a methodological contribution. The paper follows the discrete-adjoint algebra of Reuber et al. (2018), while automatic differentiation, sparse colouring and adjoint software are already available in several numerical frameworks. The manuscript needs to state whether its contribution is a new method, a reusable implementation, a set of verified examples or a pedagogical exposition. At present it makes elements of all four claims but does not establish any of them fully.

The repeated statement that the gradient has a cost independent of the number of parameters is also incomplete. One transposed linear solve is independent of the parameter dimension. Assembly of B with forward-mode automatic differentiation is not automatically so. Its cost depends on the number and sparsity structure of the parameter directions. The capsule, geothermal and plume examples use sparsity detection and colouring, whereas the shear-band example forms a dense ForwardDiff Jacobian over the parameter columns. No timings, memory measurements, colour counts or solver information are reported. For the geothermal example, which has approximately one million state unk-

nouns and several million parameter values, the reader cannot tell whether the calculation is practical.

The current design explicitly assembles J and B , uses a sparse direct factorisation, and omits the reverse traversal required by a transient problem. It is consequently restricted to two-dimensional demonstrations that fit this solver strategy. Matrix-free products, a scalable iterative solver, a preconditioner for the transposed system and checkpointing are not minor extensions of the present implementation. They are central parts of a different scalable design. The authors do not need to implement all of these in a demonstration paper, but they must bound their claim: either demonstrate a credible scalable route on one example or state that scaling to three-dimensional and long-running time-dependent problems is outside the contribution.

A revised paper would also benefit from a direct comparison with existing approaches, including the LaMEM adjoint of Reuber et al. (2018, 2020), PETSc TSAdjoint, dolfin-adjoint and G-ADOPT, and the Enzyme-based approach of Utkin et al. (2025). Such a comparison would clarify what the Julia implementation offers beyond already established methods.

The examples are not internally consistent enough to support physical interpretation

The most serious case is the geothermal model. The stratigraphic order contradicts the cited regional geology, the wells do not occupy the units stated in the text, and three different thermal boundary conditions appear in the manuscript and code. The text invokes buoyancy-driven circulation, but the fluid density is constant and no thermal expansivity occurs in the governing equations or implementation. The wells are described as pressure-controlled although the code imposes rates. The plotted Darcy flux implies negligible transport over one hundred years, the isotherms show no advected plume from the injector, and the strongest cold anomaly occurs at the producer. These features are inconsistent with the physical interpretation in the text and require a mass-balance calculation, near-well fluxes and a production-temperature history before sensitivity kernels can be interpreted.

The geothermal objective is also not the extracted power described in the manuscript. It uses absolute temperature in degrees Celsius, has unclear dimensional factors, and differs between the equation and the adjoint right-hand side. More seriously, the forward residual used for J and the separately copied residual used for B contain opposite signs for the well terms. Thus J and B are derivatives of different residual functions precisely

where the objective is evaluated. This contradicts the main premise of differentiating one residual implementation.

The plume and shear-band examples contain comparable interpretive problems. In Figure 8, the rows are described in the reverse order from their labels. The objective differentiated in the plume solver is not the dimensional, normalised quantity printed on the figure, and the plotting script applies the dimensional velocity scale a second time. The surface objective is a signed sum of vertical velocity across a closed free-slip domain and is nearly zero by mass conservation, so it is not a useful proxy for geodetic uplift. In the shear-band model, material advection is disabled. The inclusion therefore remains circular under large imposed strain, the band width is controlled by the grid and the final state is dominated by interaction with the boundaries. These simplifications can form a numerical test, but they do not support the geological interpretation presently attached to them.

What a revised manuscript would need to show

The paper could become useful if it is rebuilt around a smaller set of claims that can actually be demonstrated. The authors need to define the transient residual and distinguish a frozen-step derivative from a trajectory gradient (if the whole selling point is that one adjoint block can be calculated, the contribution is very thin). Every retained example needs a forward benchmark and a derivative test at the resolution shown. The plume adjoint and the geothermal residual must be corrected before the figures are recomputed. At least one example should report the wall-clock cost of the forward solve, assembly of B and adjoint solve. The manuscript must then compare the achieved cost with the theoretical claim (to show that this is useful you need to show one adjoint is close to one fwd cost) and state the limits of the assembled, direct-solver approach (How does the machinery of using the same solve strategies for the forward solve on the adjoint and forward run of the tape work? Do you even do this?).

The physical examples also need to be reconstructed from the implementation outward. Each equation, parameter, boundary condition, objective, plotted value and caption must describe the calculation that actually produced the result. Only after that reconciliation can the sensitivity fields support physical conclusions. I would encourage the authors to retain fewer examples if this permits each one to be verified properly.