The Journal of Open Source Software

DPFEHM: a differentiable subsurface physics simulator

Daniel O'Malley [1¶], **Sarah Y. Greer** [1,2], **Aleksandra Pachalieva** [1], **Wu Hao** [1], **Dylan Harp** [3], **and Velimir V. Vesselinov** [4]

1 Los Alamos National Laboratory, USA **2** Massachussets Institute of Technology, USA **3** The Freshwater Trust, USA **4** SmartTensors, LLC, USA ¶ Corresponding author

DOI: 10.21105/joss.04560

Software

- Review
- Repository
- Archive

Editor: Jed Brown

Reviewers:

- @WilkAndy
- @rtmills
- @rezgarshakeri

Submitted: 01 July 2022
Published: 13 October 2023

Summary

The Earth's subsurface is a key resource that provides energy via fossil fuels and geothermal sources, stores drinking water, and is used in the fight against climate change via carbon sequestration. Simulating the physical processes that occur in the Earth's subsurface with computers enables better use of this resource. `DPFEHM` is a Julia package that includes computer models with a focus on the Earth's subsurface, especially fluid flow, which is critical for the aforementioned applications. `DPFEHM` is able to solve the groundwater flow equations (single phase flow), Richards equation (air/water), the advection-dispersion equation, and the 2d wave equation. One of the key features of `DPFEHM` is that it supports automatic differentiation, so it can be integrated into machine learning workflows using frameworks such as Flux or PyTorch. The automatic differentiation capabilities give it the same performance as adjoint methods.

Statement of need

Numerical models of subsurface flow and transport such as MODFLOW (Harbaugh, 2005), FEHM (Zyvoloski et al., 1997), PFLOTRAN (Lichtner et al., 2015), etc. are ubiquitous. However, it is challenging to integrate these with machine learning frameworks for highly parameterized problems, which are common in the subsurface. Highly parameterized problems require reverse-mode differentiation (i.e., adjoint methods), which depend upon Jacobian matrices. DPFEHM relies uses automatic differentiation to compute these sparse Jacobians, but non-differentiable alternatives require significant developer time to implement these Jacobians. An automatically-differentiable model like `DPFEHM` can be seamlessly integrated into these machine learning workflows via reverse-mode differentiation. This enables machine learning workflows with `DPFEHM` in the loop, e.g., to learn to manage pressure in a scenario where wastewater or carbon dioxide are being injected into the subsurface (Pachalieva et al., 2022). For example, without automatic differentiation the machine learning would get stuck when it needed to compute a Jacobian-vector product involving the subsurface simulator. `DPFEHM` fills this gap and enables the efficient computation of the Jacobian-vector product. It is additionally useful for non-machine learning workflows, because gradient calculations are also ubiqitous in more traditional workflows such as inverse analysis (Wu et al., 2022) and uncertainty quantification (Gelman et al., 2015) (UQ). For example, inverse analysis often uses the gradient to perform some variation of gradient descent to find the solution to the inverse problem, so making this fast is important in this context. Traditional inverse modeling and UQ tools (`PEST` (Doherty & Hunt, 2010) being the most widely used example), take a non-intrusive approach, which allows them to work with any simulator but forces them to treat the simulator as a black box. `DPFEHM` lays the groundwork for next-generation UQ tools that utilize the gradient and Jacobian information that `DPFEHM` efficiently provides. Of course, it can also be used to efficiently simulate complex physics related to flow and transport in the subsurface (Greer et al., 2022) without exploiting the differentiability very deeply.

An alternative to a differentiable numerical model is to use a differentiable machine learning model that is trained on data from a non-differentiable numerical model such as those listed above. However, this approach has two major drawbacks. First, such a machine learning model may be insufficiently trustworthy, depending on the application. By contrast, `DPFEHM` uses a finite volume method that ensures local mass conservation, which is backed by rigorous accuracy guarantees. Second, in scenarios with complex physics, it may be impossible to generate sufficient data to train an adequate machine learning model. By contrast, `DPFEHM` does not require any training – just a description of the equations, boundary conditions, etc. that define the problem. Physics-based alternative models such as FEHM (Zyvoloski et al., 1997) and PFLOTRAN (Lichtner et al., 2015) currently support a wider range of physics processes, but lack automatic differentiation.

`DPFEHM` was designed to be a research tool to explore the interface between numerical models and machine learning models. To date, it has been used in several publications including (Greer et al., 2022; Pachalieva et al., 2022; Wu et al., 2022). `DPFEHM` uses a two-point flux approximation finite volume scheme, which is commonly used by commercial reservoir simulation codes. This means that an orthogonal grid is required to ensure convergence, similar to other codes such as FEHM and PFLOTRAN. Alternative codes such as Amanzi-ATS (Mercer-Smith, 2020) use a more advanced mimetic finite difference discretization that supports non-K-orthogonal grids. Other methods for non-K-orthogonal grids include multipoint flux approximations, least-squares gradient reconstruction, and mixed finite element approaches. The performance advantage for `DPFEHM` over non-differentiable alternatives such as those mentioned previously comes in computing the gradient of a function that involves the solution of a subsurface physics equation. In these settings, the cost of computing a gradient with `DPFEHM` is typically around the cost of running two physics simulations. For models that lack discrete adjoints, which covers most non-differentiable models, the cost is equal to performing a number of simulations that is proportional to the number of parameters – exorbitant when the number of parameters is large. This is important for subsurface physics, because there is often one or more parameters at each node in the mesh.

Installation

`DPFEHM` can be installed from within Julia by running `import Pkg; Pkg.add("DPFEHM")`. The installation can subsequently be tested by running `Pkg.test("DPFEHM")`.

Acknowledgements

This project was supported by Los Alamos National Laboratory LDRD project 20200575ECR. S. Y. Greer acknowledges support from the United States Department of Energy through the Computational Science Graduate Fellowship (DOE CSGF) under grant number DE-SC0019323.

References

Doherty, J. E., & Hunt, R. J. (2010). *Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration* (Vol. 2010). US Department of the Interior, US Geological Survey Reston, VA, USA. https://doi.org/10.3133/sir20105169

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, *40*(5), 530–543. https://doi.org/10.3102/1076998615606113

Greer, S., Hyman, J., & O'Malley, D. (2022). A comparison of linear solvers for resolving flow in three-dimensional discrete fracture networks. *Water Resources Research*, *58*(4), e2021WR031188. https://doi.org/10.1029/2021wr031188

Harbaugh, A. W. (2005). *MODFLOW-2005, the US geological survey modular ground-water model: The ground-water flow process* (Vol. 6). US Department of the Interior, US Geological Survey Reston, VA, USA. https://doi.org/10.5066/F7RF5S7G

Lichtner, P. C., Hammond, G. E., Lu, C., Karra, S., Bisht, G., Andre, B., Mills, R., & Kumar, J. (2015). *PFLOTRAN user manual: A massively parallel reactive flow and transport model for describing surface and subsurface processes*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States); Sandia …. https://doi.org/10.2172/1168703

Mercer-Smith, J. A. (2020). *Amanzi–ATS: Modeling environmental systems across scales [brief]* (LA-UR-20-26636). Los Alamos National Lab.(LANL), Los Alamos, NM (United States). https://doi.org/10.2172/1657092

Pachalieva, A., O'Malley, D., Harp, D. R., & Viswanathan, H. (2022). Physics-informed machine learning with differentiable programming for heterogeneous underground reservoir pressure management. *Scientific Reports*, *12*(1), 18734. https://doi.org/10.1038/s41598-022-22832-7

Wu, H., O'Malley, D., Golden, J. K., & Vesselinov, V. V. (2022). Inverse analysis with variational autoencoders: A comparison of shallow and deep networks. *Journal of Machine Learning for Modeling and Computing*, *3*(2). https://doi.org/10.1615/jmachlearnmodelcomput.2022042093

Zyvoloski, G. A., Robinson, B. A., Dash, Z. V., & Trease, L. L. (1997). *Summary of the models and methods for the FEHM application-a finite-element heat-and mass-transfer code*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States). https://doi.org/10.2172/14903