An Exponential Time Integration Perspective on Machine Learning with Neural Networks

Martin Schreiber*

Time integration of PDEs to simulate the atmosphere is a non-trivial task once considering wallclock time as well as accuracy constraints. E.g. solving long-term simulations such as Paleo climate simulations has a wallclock time of weeks or even months and weather simulations must be completed within a time frame of about an hour to ensure the results to be included in the daily weather forecast. Overcoming these restrictions requires interdisciplinary research: In high-performance computing, for example, limitations of stagnating speed of processors, increase in parallel running cores as well as vector lengths and communication overheads must be coped with. On the mathematics side, time integration methods must be developed that not only offer larger time step sizes, but also ones which lead to a reduced wallclock-time vs. error rate through the inclusion of mathematics and high-performance computing.

A promising approach is the exponential integration method (related and partly identical to time integration with Laplace transforms and semi-groups via the Cauchy contour integral method) which lifts the time step size restrictions for stiff linear PDEs. In the context of dynamical cores, this method has been investigated over the last few years with the shallow-water equations on the rotating sphere for linear [4] and non-linear [1, 5] formulations with the linear(ized) exponential integration itself based on rational approximations with Fourier (T-REXI [2]), Cauchy Contour integration (CI-REXI [5]) and Krylov subspace methods [1]. Recent research also revealed a reduction of errors, once combining exponential integrators with semi-Lagrangian methods (SL-EXP [3]) for the time integration of the non-linear advection part. However, an increase in maximum stable time step size was not achieved. This once again illustrates the importance of time step size limitations induced by the non-linear interactions and motivates further research to overcome them.

There are claims about machine learning with neural networks to be able to approximate high-dimensional spaces and also non-linearities due to the activation functions which motivates an investigation of their suitability to overcome time step restrictions in the context of dynamical cores. Considering existing work based on neural networks, time integrating PDEs is often conducted in a black box fashion, taking an off-the-shelf neural network and training it with RMS (L_2-based) error rates larger than $O(10^{-3})$ and also entirely ignoring L_∞ norms which are typically even larger. Such high error rates make these trained networks clearly insufficient for a stable time integration of atmospheric equations, raising the task to improve for this.

This work picks up this thought of exploiting neural networks for temporal integration of PDEs, targeting large time step sizes beyond the CFL condition. We design the underlying neural network with an inspiration from exponential integrators: Instead of using arbitrary neural networks in a blackbox fashion, we will design neural networks to reflect the underlying linear terms and non-linear interactions.

First, purely linear one-dimensional problems will be discussed and the aforementioned problems of high error of state-of-the-art gradient-based optimizers which interestingly already exists for linear optimization of neural networks. As a first step towards a robust training, we will overcome this with a reformulation to a linear optimization problem and a conjugate gradient optimizer. As a second part, possible extensions to non-linear problems will be discussed which all target a low-error prediction. Regarding high-performance computing and taking into account that neural networks are equipped with a high potential of parallelization with their tensor-dot product structure, such methods might become interesting for future FPGA-based and neural-network-oriented architectures.

References

[1] Stéphane Gaudreault and Janusz A. Pudykiewicz. An efficient exponential time integration method for the numerical solution of the shallow water equations on the sphere. *Journal of Computational Physics*, 322:827–848, 2016.

*Technical University of Munich, martin.schreiber@tum.de

[2] T. S. Haut, T. Babb, P. G. Martinsson, and B. A. Wingate. A high-order time-parallel scheme for solving wave propagation problems via the direct construction of an approximate time-evolution operator. *IMA Journal of Numerical Analysis*, 36(2):688–716, 2016.

[3] P S Peixoto and M Schreiber. Semi-Lagrangian Exponential Integration With Application To the Rotating Shallow Water Equations. *SIAM Journal on Scientific Computing*, 2019.

[4] Martin Schreiber and Richard Loft. A parallel time integrator for solving the linearized shallow water equations on the rotating sphere. *Numerical Linear Algebra with Applications*, 26(2), 2018.

[5] Martin Schreiber, Nathanaël Schaeffer, and Richard Loft. Exponential integrators with parallel-in-time rational approximations for the shallow-water equations on the rotating sphere. *Parallel Computing*, 85:56–65, 2019.