The first AI simulation of a black hole

Rodrigo Nemmen[1] , Roberta Duarte[1] and João P. Navarro[2]

[1]Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas,
Departamento de Astronomia, São Paulo, SP 05508-090, Brazil
email: rodrigo.nemmen@iag.usp.br

[2]NVIDIA

Abstract. We report the results from our ongoing pilot investigation of the use of deep learning techniques for forecasting the state of turbulent flows onto black holes. Deep neural networks seem to learn well black hole accretion physics and evolve the accretion flow orders of magnitude faster than traditional numerical solvers, while maintaining a reasonable accuracy for a long time.

Keywords. Black hole physics, astrostatistics, active galactic nuclei

1. Introduction

My presentation was supposed to be about the new constraints on the spin of the supermassive black hole (SMBH) in M87. Here is the bottom line from that work: we have constrained the spin parameter to be $|a_*| > 0.4$ (Nemmen 2019). The spin is the second fundamental parameter of black hole (BH) spacetimes. This constraint should set expectations for future estimates of the M87* spin with the Event Horizon Telescope and other observatories.

Instead, I will present some early exciting results from our pilot investigation of artificial intelligence (AI) methods as tools to accelerate numerical simulations of BH accretion flows. Here, we address two inter-related questions: Can we make the models faster while maintaining an accuracy comparable to explicit solvers of the fluid conservation equations? Can deep neural networks learn fluid dynamics?

2. Deep learning

Let me begin with the fundamental problem of BH astrophysics: to figure out the function

$$\text{AGN}(t) = f(M, a_*, \dot{M}) \tag{2.1}$$

which quantifies the complete time-evolution of "weather" around SMBHs, where M and $\dot{M}$ are the BH mass and mass accretion rate, respectively. The challenge is that BH weather is a complex process, requiring the solution of nonlinear, multidimensional partial differential equations which are very time-consuming (e.g. Porth *et al.* 2019). Ideally, we want those simulations to have a duration much shorter than a typical PhD thesis timescale, so we need to make them as fast as possible.

Here, we are investigating the use of deep learning (DL) techniques for that purpose. DL consists of using deep neural networks inspired by the way the brain works, with a large number of layers and parameters ("neurons") (Goodfellow, Bengio & Courville 2016). Deep neural networks are good approximators for empirical functions which are

too complex to be have an analytical form (e.g. Cybenko 1989). It is not an exaggeration to say that a considerable fraction of AI work today consists of applications—and improvements upon—DL.

In practice, DL algorithms "learn from experience": instead of explicitly coding the instructions in the code, one trains the machine by showing a lot of examples and comparing the output of trained algorithm to a test dataset. Then, by tweaking the network architecture and its hyperparameters, one arrives at a trained deep net (LeCun 2015). DL is leading to several breakthroughs in many fields (e.g. Mnih *et al.* 2015; Silver *et al.* 2016; Krizhevsky, Sutskever & Hinton 2017), including astronomy (e.g. Hausen & Robertson 2020; Zhang *et al.* 2019). The downsides of DL is that it is data-hungry (needs a lot of data to be effective) and computationally expensive to train. Once trained, however, the algorithm deploys answers very quickly.

Here, we devise the prediction challenge as a computer vision problem and infer the evolution of the system from the sequence of input data cubes that comprise previous states of an accreting BH. This is a data-driven, equation-free approach in which a model learns to approximate the relevant physics from the training examples alone and not by incorporating a priori knowledge about the equations underlying the processes. This is similar to the work of Jaeger & Haas (2004); Tompson *et al.* (2016).

3. Teaching a machine about BH accretion

The training data set consists of the hydrodynamical BH accretion simulations performed by Almeida & Nemmen (2020). This is a set of long-duration (durations $>10^5 \, GM/c^3$), 2D models designed to explore the winds produced by low-luminosity active galactic nuclei, where a Schwarzschild BH is fed by a hot, geometrically thick accretion flow. The specific model we chose is PNSS3, which has one of the longest durations among the models computed by Almeida & Nemmen (2020). The numerical simulation computes the spacetime distribution of gas densities around the BH which we feed to the DL model.

For the learning algorithm, we use the well-known U-Net convolutional neural network (ConvNet) architecture, which is commonly used to extract information from datasets which involve spatial and temporal coherence (e.g. Karpathy *et al.* 2014). For the training, we divide the time series data into 67.5% training, 12.5% cross-validation and 20% testing phases. How well does the trained DL model predict the future state of the density field around the BH?

4. Preliminary results

We quantify the performance of the DL approach by comparing a number of indicators with those obtained from the explicit solution to the conservation equations (Duarte, Nemmen & Navarro, in preparation). Here, we present some early, preliminary results.

The left panel of Figure 1 displays the density difference between the numerical simulations (i.e. the target) and the trained ConvNet (the prediction) as a function of time. What we mean by error here is defined in the lower left corner of Figure 1. We see that the error gradually builds up over time, as if the predictions drifted from the ground truth solutions of the conservation equations. The middle panel shows the target density map, which results from solving the fluid conservation equations. The right panel corresponds to the inferred deep learning model—the neural network's imagination.

Figure 2 summarizes the main result of this presentation. Each panel displays the density marginalized over the polar angle. Think of this as the average density in spherical shells of increasing radii. On a first inspection, we can see that the DL forecast and the data are virtually identical. Only after analyzing the residuals we realize that
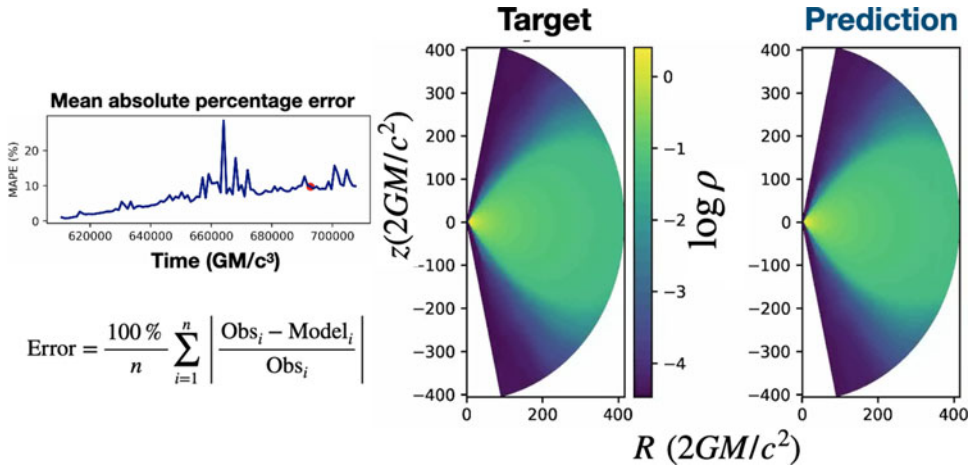
Figure 1. Comparison between the deep neural network's "imagination" (indicated as *prediction*) and the actual numerical solutions from the fluid conservation laws (the *target*), for a black hole surrounded by a hot accretion flow. Left panel: time evolution of the difference between the prediction and target (i.e. the error of the trained model). The lengths are expressed in terms of the Schwarzschild radius. Time in gravitational units.
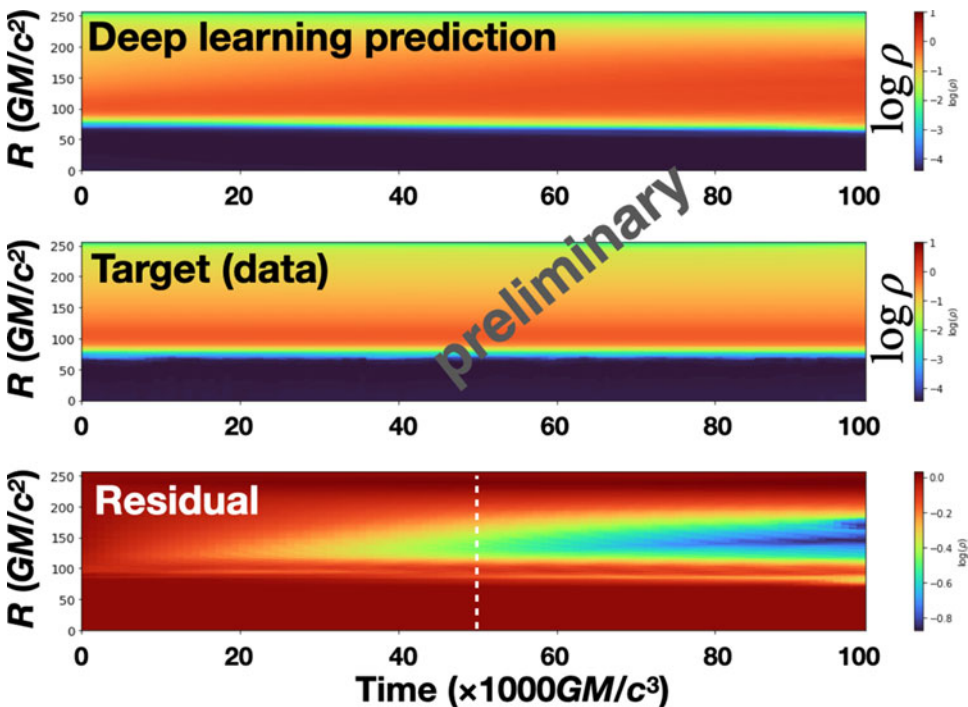
Figure 2. Prediction of deep learning model compared with the data from the numerical simulations of BH accretion. The colors map the logarithm of the gas density. The density is averaged over the polar angle. The residuals are the difference between the logarithms. The vertical white line indicates the time at which the deep nets begin to drift away from the ground truth.

there are differences between the two. We see that the neural network imagines the future well for a duration of about $5 \times 10^4 GM/c^3 \approx 33t_{\mathrm{dyn}}$, where t_{dyn} is the dynamical time at 100 Schwarzschild radii. Relatively speaking, this is a long time considering the timescales that regulate the system. Only after $33t_{\mathrm{dyn}}$ the neural network begins displaying symptoms of an "artificial Alzheimer's disease".

These results indicate that DL techniques are promising for forecasting the state of multidimensional chaotic systems. While the trained model is performing well—i.e. reproducing the target data accurately—it does so much faster than the original simulation: about *600 times faster*. The DL training was done using two NVIDIA GPUs ($\approx$40 TFlops, FP32); the hydrodynamic simulation was performed on a CPU cluster using 200 cores in parallel ($\sim$3 TFlops, FP64). Even taking into account the different performances of the processors used, the DL model is still vastly faster than the original modeling approach used to generate the training data.

5. Summary

This work is a pilot study of the performance of deep learning in forecasting the state of large spatiotemporally chaotic systems comprised by turbulent flows onto black holes. So far, we have obtained promising preliminary results. Not only deep neural networks seem to learn well black hole accretion physics: they evolve the accretion flow $\sim$1000 times faster than traditional numerical solvers. The future seems bright for black hole numerical investigations, in which science will be less restricted by hardware limitations.

References

Almeida, I. & Nemmen, R. 2020, *MNRAS*, 492, 2553
Cybenko, G. 1989, *Math. of Cont., Sign. and Sys.*
Goodfellow, I., Bengio, Y., & Courville, A. 2016, *The MIT Press*
Hausen, R. & Robertson, B. E. 2020, *ApJS*, 60, 84
Jaeger, H. & Haas, H. 2004, *Science*, 304, 78
Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. 2014 *CVPR*
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2017, *Commun. ACM*, 60, 84
LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436
Mnih, V., Kavukcuoglu, K., Silver, D., *et al.* 2015, *Nature*, 518, 529
Nemmen, R. 2019, *ApJL*, 880, L26
Porth, O. & others 2019, *ApJS*, 243, 26
Silver, D., Huang, A., Maddison, C. J., *et al.* 2016, *Nature*, 529, 484
Tompson, J., Schlachter, K., Sprechmann, P., *et al.* 2016, arXiv:1607.03597
Zhang, X., Wang, Y., Zhang, W., *et al.* 2019, arXiv:1902.05965

Discussion

WEINBERGER: From the movie you showed us, it seems that the accretion flow is almost stationary. I do not see much turbulence going on.

NEMMEN: We have chosen for the training data a model from Almeida & Nemmen (2020) that shows little variability, because this was the longest model we had (more data improves the performance of the training). We are also investigating and generating data with much more variability and comparable durations.

SANCHEZ: I heard that there are also some GPU codes that solve fluid dynamics and are faster than CPU solvers. Did you compare the speedups of your deep learning approach with those GPU solvers?

NEMMEN: I only know of one GPU solver for black hole accretion simulations, H-AMR developed by Liska et al. In my understanding, H-AMR is obtaining speedups of about 10x when comparing GPUs and multicore CPUs. Our early results indicate a speedup of about 1000x using a good GPU of the kind that gamers use. So it seems that a DL approach would offer a great cost-benefit. But much remains to be investigated yet.

WYLEZALEK: Can you trust the predictions of an algorithm that does not solve the physical laws? In other words, can you really do physics without doing physics? You don't know what the AI algorithm is really doing, how it arrives at a given answer.

NEMMEN: This is a big issue in AI nowadays: the issue of explainability—explaining how the neural networks arrive at a given output. I do not have an answer for that. But I do know that there is progress in this front by some AI research groups. Using AI for the purpose described in this presentation involves a shift in the underlying philosophy of the numerical simulations. What if it does indeed work well for forecasting while being much faster than current simulations? I think this is a very interesting question.