Consistent Energy Preserving Neural Networks

Shaan A. Desai[1,2] **David Sondak**[2] **Marios Mathiakis**[2]
Machine Learning Research Group
University of Oxford [1]
Institute of Applied Computational Science
Harvard University [2]

Abstract

Unlike vanilla Artificial Neural Networks, Neural Networks embedded with physically-informed priors achieve remarkable results in accurately learning and predicting non-linear dynamical systems. Despite this success, their generalization performance is often limited to short-range trajectories that run for less than the final training time. The performance degradation is due to state and energy drift induced by errors accumulating over time and it poses a critical challenge to learning highly non-linear and chaotic long-range dynamics precisely when data for only a limited number of short-range trajectories are available. One proposed solution to this challenge has been to incorporate symplectic integrators that aim to preserve the symplectic flow. Here, we present an alternative, an L1 penalty of the Hamiltonian derivative with respect to time: $d\mathcal{H}(p, q, t)/dt$. The inclusion of this term, a straightforward addition to the network, enforces the Hamiltonian to remain constant in time for a given trajectory. We empirically illustrate the stabilising effect this penalty has on our network. In addition, we use it solve problems in non-linear dynamics such as heinon-heiles, 3-body problem and double pendulum.

1 Introduction

- Physicists have long been interested in learning the governing equations of complex dynamical systems. - Neural networks, as universal function approximators, have shown resounding success across a host of domains. - Naturally, physicists decided to use neural networks to learn dynamical systems. - However, their performance in learning physical systems has often been limited. - Physicists, as a consequence, have not been encouraged by the initial excitement of machine learning. - However, new research aimed at *scientific machine learning* - a branch that tackles science problems with domain-specific ML, is paving a way to address numerous challenges. - One crucial method has been to incorporate prior theoretical information into the network, such as hamiltonian mechanics. - This excitement has spurred others to work with lagrangians, ODEs and even graphs in order to tackle dyanamical systems. - Despite their widespread adoption, Hamiltonian Neural Networks still struggle to predict long-range trajectories precisely. This is in part due to the energy drift that ensues as we accumulate errors over time. - A major topic of discussion to address this issue has been to adopt symplectic integrators that preserve the symplectic flow. - However, we identify an additional component, often neglected when stating Hamilton's equations, that significantly stabilises learning. - We show that the addition of the dh/dt penalty induces the network to conserve energy across time for a given trajectory, a facet not yet invoked by neural networks. - We extensively benchmark this addition across multiple datasets and consistently find the inclusion to be of benefit. - Furthermore, we emphasise that the constraint is an easy plug-and-play addition to existing networks and illustrate how existing networks such as HNN, Symp ODEN and Hnets benefit from its inclusion. - Most importantly, we take the HNN model used for Heinon-Heiles and include a symplectic integrator with dhdt.

Proceedings of the 24$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

2 Background

2.1 Hamiltonian Neural Networks

Recently, [?] demonstrated that dynamic predictions through time can be improved using Hamiltonian Neural Networks (HNNs) which endow models with a Hamiltonian constraint. The Hamiltonian is an important representation of a dynamical system because it is one of two approaches that generalizes classical mechanics. The Hamiltonian $\mathcal{H}$ is a scalar function of position $\mathbf{q} = (q_1, q_2, ..., q_M)$ and momentum $\mathbf{p} = (p_1, p_2, ..., p_M)$. In representing physical systems with a Hamiltonian, one can simply extract the time derivatives of the inputs by differentiating the Hamiltonian with respect to its inputs (see Eqn. 1.)

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} \tag{1}$$

As a consequence, it is noted in [**?**] that by accurately learning a Hamiltonian, the system's dynamics can be naturally extracted through backpropagation. This information allows us to build two 1st-order differential equations which can be used to update the state space, $(\mathbf{q}, \mathbf{p})$. Equation 2 shows this integral, in which we define the symplectic gradient $\mathbf{S} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}}, -\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\right]$:

$$(\mathbf{q}, \mathbf{p})_{t+1} = (\mathbf{q}, \mathbf{p})_t + \int_t^{t+1} \mathbf{S}(\mathbf{q}, \mathbf{p}) \mathrm{d}t \tag{2}$$

It can be shown that the Hamiltonian in many systems also represents the total energy of the system. Therefore, the Hamiltonian is a powerful inductive bias that can be utilised to evolve a physical state while maintaining energy conservation.

3 Method

The time-derivative of a Hamiltonian $\mathcal{H}(\mathbf{q}, \mathbf{p}, t)$ can be obtained using the chain rule:

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{q}} \frac{\partial \mathbf{q}}{\partial t} + \frac{\partial H}{\partial \mathbf{p}} \frac{\partial \mathbf{p}}{\partial t} + \frac{\partial \mathcal{H}}{\partial t} \tag{3}$$

The underlying fact in energy preserving systems is that $H = E$ is a constant. As such, the time derivative should be set to zero. By learning the Hamiltonian and differentiating it with respect to $\mathbf{q}, \mathbf{p}$ we obtain the time derivatives of the state vectors which when replaced in the above equation yield:

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = -\frac{\partial \mathbf{p}}{\partial t} \frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{q}}{\partial t} \frac{\partial \mathbf{p}}{\partial t} + \frac{\partial \mathcal{H}}{\partial t} = 0 \tag{4}$$

which leaves us with:

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial t} = 0 \tag{5}$$

To enforce this additional constraint in our networks, all we need to do is simply provide the time to the network, which extends the input dimension by 1, and allow the time component to be differentiable. We then compute the gradient and use an L1 penalty such that the final loss is:

$$\mathcal{L}_{EPNN} = \left\| \frac{\partial \mathcal{H}_\theta}{\partial \mathbf{q}} + \frac{\partial \mathbf{p}}{\partial t} \right\| + \left\| \frac{\partial \mathcal{H}}{\partial \mathbf{p}} - \frac{\partial \mathbf{q}}{\partial t} \right\| + \left| \frac{\partial \mathcal{H}}{\partial t} \right| \tag{6}$$

4 FIRST LEVEL HEADINGS

First level headings are all caps, flush left, bold, and in point size 12. Use one line space before the first level heading and one-half line space after the first level heading.

4.1 Second Level Heading

Second level headings are initial caps, flush left, bold, and in point size 10. Use one line space before the second level heading and one-half line space after the second level heading.

4.1.1 Third Level Heading

Third level headings are flush left, initial caps, bold, and in point size 10. Use one line space before the third level heading and one-half line space after the third level heading.

Fourth Level Heading Fourth level headings must be flush left, initial caps, bold, and Roman type. Use one line space before the fourth level heading, and place the section text immediately after the heading with no line break, but an 11 point horizontal space.

4.2 Citations, Figure, References

4.2.1 Citations in Text

Citations within the text should include the author's last name and year, e.g., (Cheesman, 1985). Be sure that the sentence reads correctly if the citation is deleted: e.g., instead of "As described by (Cheesman, 1985), we first frobulate the widgets," write "As described by Cheesman (1985), we first frobulate the widgets."

The references listed at the end of the paper can follow any style as long as it is used consistently.

4.2.2 Footnotes

Indicate footnotes with a number[1] in the text. Use 8 point type for footnotes. Place the footnotes at the bottom of the column in which their markers appear, continuing to the next column if required. Precede the

[1]Sample of the first footnote.

footnote section of a column with a 0.5 point horizontal rule 1 inch (6 picas) long.[2]

4.2.3 Figures

All artwork must be centered, neat, clean, and legible. All lines should be very dark for purposes of reproduction, and art work should not be hand-drawn. Figures may appear at the top of a column, at the top of a page spanning multiple columns, inline within a column, or with text wrapped around them, but the figure number and caption always appear immediately below the figure. Leave 2 line spaces between the figure and the caption. The figure caption is initial caps and each figure should be numbered consecutively.

Make sure that the figure caption does not get separated from the figure. Leave extra white space at the bottom of the page rather than splitting the figure and figure caption.

This figure intentionally left non-blank

Figure 1: Sample Figure Caption

4.2.4 Tables

All tables must be centered, neat, clean, and legible. Do not use hand-drawn tables. Table number and title always appear above the table. See Table 1.

Use one line space before the table title, one line space after the table title, and one line space after the table. The table title must be initial caps and each table numbered consecutively.

Table 1: Sample Table Title

PART	**DESCRIPTION**
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

5 SUPPLEMENTARY MATERIAL

If you need to include additional appendices during submission, you can include them in the supplementary material file. You can submit a single file of additional supplementary material which may be either a pdf file (such as proof details) or a zip file for other formats/more files (such as code or videos). Note that reviewers are under no obligation to examine your supplementary material. If you have only one supplementary pdf file, please upload it as is; otherwise gather everything to the single zip file.

You must use `aistats2021.sty` as a style file for your supplementary pdf file and follow the same formatting instructions as in the main paper. The only difference is that it must be in a *single-column* format. You can use `supplement.tex` in our starter pack as a starting point. Alternatively, you may append the supplementary content to the main paper and split the final PDF into two separate files.

6 SUBMISSION INSTRUCTIONS

To submit your paper to AISTATS 2021, please follow these instructions.

1. Download `aistats2021.sty`, `fancyhdr.sty`, and `sample_paper.tex` provided in our starter pack. Please, do not modify the style files as this might result in a formatting violation.
2. Use `sample_paper.tex` as a starting point.
3. Begin your document with

   ```
   \documentclass[twoside]{article}
   \usepackage{aistats2021}
   ```

   The `twoside` option for the class article allows the package `fancyhdr.sty` to include headings for even and odd numbered pages.
4. When you are ready to submit the manuscript, compile the latex file to obtain the pdf file.
5. Check that the content of your submission, *excluding* references, is limited to **8 pages**. The number of pages containing references alone is not limited.
6. Upload the PDF file along with other supplementary material files to the CMT website.

6.1 Camera-ready Papers

If your papers are accepted, you will need to submit the camera-ready version. Please make sure that you follow these instructions:

1. Change the beginning of your document to

   ```
   \documentclass[twoside]{article}
   \usepackage[accepted]{aistats2021}
   ```

[2]Sample of the second footnote.

The option `accepted` for the package `aistats2021.sty` will write a copyright notice at the end of the first column of the first page. This option will also print headings for the paper. For the *even* pages, the title of the paper will be used as heading and for *odd* pages the author names will be used as heading. If the title of the paper is too long or the number of authors is too large, the style will print a warning message as heading. If this happens additional commands can be used to place as headings shorter versions of the title and the author names. This is explained in the next point.

2. If you get warning messages as described above, then immediately after `\begin{document}`, write

```
\runningtitle{Provide here an alternative
shorter version of the title of your
paper}
\runningauthor{Provide here the surnames
of the authors of your paper, all
separated by commas}
```

   Note that the text that appears as argument in `\runningtitle` will be printed as a heading in the *even* pages. The text that appears as argument in `\runningauthor` will be printed as a heading in the *odd* pages. If even the author surnames do not fit, it is acceptable to give a subset of author names followed by "et al."

3. The camera-ready versions of the accepted papers are 8 pages, plus any additional pages needed for references.

4. If you need to include additional appendices, you can include them in the supplementary material file.

5. Please, do not change the layout given by the above instructions and by the style file.

Acknowledgements

All acknowledgments go at the end of the paper, including thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support. To preserve the anonymity, please include acknowledgments *only* in the camera-ready papers.

References

References follow the acknowledgements. Use an unnumbered third level heading for the references section. Please use the same font size for references as for the body of the paper—remember that references do not count against your page length total.

References

J. Alspector, B. Gupta, and R. B. Allen (1989). Performance of a stochastic learning microchip. In D. S. Touretzky (ed.), *Advances in Neural Information Processing Systems 1*, 748–760. San Mateo, Calif.: Morgan Kaufmann.

F. Rosenblatt (1962). *Principles of Neurodynamics.* Washington, D.C.: Spartan Books.

G. Tesauro (1989). Neurogammon wins computer Olympiad. *Neural Computation* **1**(3):321–323.