

1 We thank the reviewers for both their favorable and their critical feedback. We have used the latter to improve and  
 2 clarify the paper. Below we group and respond to key issues, highlighting the changes we've made:

3 **1. Tasks 1-3 are perhaps too simple (R1).** We agree that tasks 1-3 are simple in that they are easy to solve with  
 4 existing methods. We included them not to showcase the real-world potential of HNNs but simply to allow for a  
 5 scientific investigation of their basic properties. Task 1 was used to show that optimizing the symplectic gradient of a  
 6 neural network is a viable and stable strategy. We then asked whether an HNN could learn a nonlinear vector field, and  
 7 Task 2 provided a sanity check that it could. We then asked whether an HNN would learn something sensible from  
 8 real-world data, and Task 3 showed that it did. Once these sanity checks were complete, we applied our approach to  
 9 the more complex and exciting Tasks 4 and 5 which we believe better showcase the power of HNNs. While we could  
 10 remove tasks 1-3 due to their simplicity, we feel the rigor and clarity are improved thanks to their inclusion.

11 **2. Hamiltonian is not generally proportional to  $q^2 + p^2$  (line 19) (R1).** This part was poorly written: we had not  
 12 intended to claim that the Hamiltonian will be generally proportional to  $q^2 + p^2$ . We've updated the text as shown in *a*  
 13 in the figure below.

14 **3. Fix issues with structure: Table 1 is before tasks 4 and 5; Figure 6 not found (R1).** We agree that these parts  
 15 were a mess. Table 1 has been moved to page 7 of the paper so that the reader can view it after reading the explanations  
 16 of Tasks 4 and 5. Table 1 was poorly formatted; we have added confidence intervals and updated it to be much more  
 17 readable. These changes are reflected in *c* in the figure below. As for Figure 6, it has been moved to the Appendix and  
 18 is now more clearly referenced as such as shown in *b* in the figure below.

19 **4. Add references for optimizing gradients of neural networks (R2)** Thanks for the five references. All are great  
 20 additions to the paper and we have added them along with a discussion of the overlap (see part *d* in the figure below).

21 **5. Show a simple control example, especially with the latent coordinate space (R3).** Great idea! We believe that  
 22 applying HNNs to control tasks will produce compelling results. However, to limit the scope of this paper to that of  
 23 introducing the core concept of an HNN, we decided to leave this application to future work.

24 **6. Model systems with (a) friction/damping or (b) contact modeling (R3).** Thanks for these suggestions. When we  
 25 were writing the paper we discussed some related experiments: how to model (a) a damped harmonic oscillator and (b)  
 26 balls bouncing in a box. Ultimately, we decided to focus this paper on a thorough and principled investigation of the  
 27 key effect and leave these more complicated scenarios to future work.

28 **7. Outline applications where the equations of motion are not sufficient/where HNNs are needed (R2). Address  
 29 significance of HNNs, especially for solving real-world problems (R1).** One key difference between HNNs and  
 30 potential energy-based approaches is that the latter generally require a reference energy obtained via electronic structure  
 31 calculations (Equation 10 in Behler 2011 and Equations 1,2 in Pukrittayakamee 2009). Meanwhile, HNNs are trained  
 32 in an unsupervised manner: we do not require reference energies. Because of this, HNNs are promising for datasets  
 33 that have unusual coordinate systems (e.g. Task 5: Pixel Pendulum) such that the reference energies are not trivial to  
 34 compute. We are very excited about real-world applications of this technique. Since releasing a preprint of this paper,  
 35 we have heard from one group that is currently running experiments with HNNs to calibrate quantum computers and  
 36 another group that is considering using HNNs to learn/calibrate the gaits of a biped robot.

**a) Rewording line 19**

directly from data. This generally presents them from learning exact physical laws. Consider the  
 discussion mass-spring system shown in Figure 1. Here the total energy of the system is being  
 conserved. More specifically, the particular coordinate system is assumed proportional to  $q^2 + p^2$ .

directly from data. This generally presents them from learning exact physical laws. Consider the  
 discussion mass-spring system shown in Figure 1. Here the total energy of the system is being  
 conserved. More specifically, the particular coordinate system is assumed proportional to  $q^2 + p^2$ .

**3 Learning a Hamiltonian from Data**

Optimizing the gradients of a neural network is a new approach. There are a few previous works  
 which do this [10, 20, 21] but their scope and implementation details diverge from this work and  
 from one another. With this in mind, our first step was to investigate the empirical properties of HNNs on  
 three simple physics tasks.

**3 Learning a Hamiltonian from Data**

Optimizing the gradients of a neural network is a new approach. There are a few previous works  
 which do this [10, 20, 21] but their scope and implementation details diverge from this work and  
 from one another. With this in mind, our first step was to investigate the empirical properties of  
 HNNs on three simple physics tasks.

**b) Figure 6 moved to the Appendix**

eventually diverge. Moreover, the HNNs converge to a quantity that already resembles total energy  
 and diverges more slowly or not at all.

HNNs and its baseline are both competitive tasks and our baseline, however, is not. The HNN consistently  
 outperforms the baseline on the energy metric (Table 1).

eventually diverge. Moreover, the HNNs converge to a quantity that already resembles total energy  
 and diverges more slowly or not at all.

HNNs and its baseline are both competitive tasks and our baseline, however, is not. The HNN consistently  
 outperforms the baseline on the energy metric (Table 1).

**Table 1. Quantitative results across all test tasks. Whichever the HNN is competitive with the baseline on  
 the test metric, it is denoted by a green background. For baseline on the energy metric, see Appendix A for a  
 description of the large test metric Table 1.**

Task Name	Baseline		HNN		Energy MSE
	Baseline	HNN	Baseline	HNN	
Task 1: Mass-spring	0.01	0.01	0.01	0.01	0.0004
Task 2: 3-body pendulum	0.01	0.01	0.01	0.01	0.002
Task 3: 3-body pendulum	0.02	0.02	0.02	0.02	0.002
Task 4: 2-body pendulum + 1D	0.02	0.02	0.02	0.02	0.009
Task 5: Pixel pendulum	0.01	0.01	0.01	0.01	0.002

**Table 1. Quantitative results across all test tasks. Whichever the HNN is competitive with the baseline on  
 the test metric, it is denoted by a green background. For baseline on the energy metric, see Appendix A for a  
 description of the large test metric Table 1.**

Task	Baseline		HNN		Energy
	Baseline	HNN	Baseline	HNN	
1: Mass-spring	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.0004 ± 0.0001
2: 3-body pendulum	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.002 ± 0.0001
3: 3-body pendulum	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.002 ± 0.0001
4: 2-body pendulum + 1D	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.009 ± 0.0001
5: Pixel pendulum	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.002 ± 0.0001

**d) Additional references:**  
 [28]; (Pukrittayakamee, 2009), [3]; (Behler, 2011), [11]; (Gastegger, 2015), [36]; (Schuett, 2017), [44]; (Yao, 2018)

**c) Move and re-format Table 1**

Changes made to paper in response to reviews