We thank the reviewers for their supportive feedback. We are glad that the reviewers found our method's performance to be "impressive" (R2), and the approach to be "very sound" (R4) and "very important in the area of learning physical systems" (R3). Our main contributions: (1) we show both analytically and empirically that using generalized coordinates to enforce constraints complicates the learning problem, and hinders generalization, (2) we propose a method that circumvents these problems by embedding into Cartesian coordinates and enforcing constraints explicitly in the learned Hamiltonian or Lagrangian dynamics, and (3) we empirically show that our approach is up to *100 times more data-efficient and up to 260 times more accurate* than previous methods, and outperforms previous methods on all systems. Our models' large gains in performance highlights the practical importance of representing the systems' dynamics in the simplest functional form possible, which has not been considered in previous works.

[R1] **"The approach, which uses regularization in Cartesian coordinates, sounds somehow straightforward."** The summary of our approach as "regularization in Cartesian coordinates" is incorrect. Our method is not regularization and we never even mention "regularization" in the paper. Unlike a regularized model, the constrained equations of motion we derive in section 5 *exactly* preserve the constraints, regardless of the learned Hamiltonian $\mathcal{H}$. In Appendix B3 we show that numerical integration of these equations preserves the constraints up to numerical tolerances. We also emphasize that in the context of outstanding results, being "straightforward" is a strength, not a weakness. Our method results in ~ 100 times more accurate and data-efficient models. Additionally, we believe working out how to embed the dynamics of complex 3D systems entirely into the evolution of Euclidean coordinates under constraints (without using angles, quaternions, etc) is also a substantial contribution.

[R1] **"The procedure is restricted to physical systems."** Indeed, our approach is restricted to systems in physical space, as we mentioned in line 285. Such systems are prevalent in engineering and robotics since our world is one such system. Our goal is to demonstrate how to learn the dynamics of such common physical systems more effectively.

[R1] **"the paper does not include several necessary information [for reproducibility]"** We detail the networks architectures in appendix D2, which are 3-layer MLPs. While we do reference this in the main text, some of the relevant details are distributed across the text, and so we will bring these together when summarizing the approach. Regarding reproducibility, our paper checks off every relevant item from the NeurIPS reproducibility checklist: we run multiple trials with error bars, release the source code, and include all hyperparameters, tuning procedures, and detailed derivations in the appendix.

[R2] **"In classical mechanics, we typically try to avoid the explicit constraints as they are numerically more brittle."** While evolving dynamics that explicitly enforce constraints numerically can lead to small accumulation of constraint drift, we quantify this drift in appendix B3, and we feel that the superior modeling ability more than makes up for this slight downside. Furthermore, the generalized coordinate approach of enforcing constraints comes with its own set of numerical instabilities, such as coordinate singularities and gimbal lock, that do not affect our proposed explicitly constrained models.

[R2,R4] **Testing on robotic systems and real world systems.** We agree this is an interesting direction for future work. We note that overall we perform a relatively exhaustive empirical evaluation, including the introduction of complex physical systems that challenge current approaches to learning Hamiltonians and Lagrangians. Developing a test setup for a robotic system is a considerable undertaking, and among the related methods of NeuralODE, HNN, DeLaN, LNN, SymODEN, SRNN and LieConv, only DeLaN was tested on real world mechanical data. In this paper we focus on simulated data which may also make it easier for others to build on our work and reproduce our results.

[R3] **"cases where the proposed parameterization is not really advantageous."** The approach is limited to systems in physical space which we briefly discussed in the conclusion. Another limitation is that the form of the constraints $\{\Phi(x)_i\}$ must be known ahead of time, such as the joint connectivity. We view the possibility of learning these constraints from data as a promising direction for future work. We will expand on these discussions.

[R3] **"What $V(X)$ denotes?"** $V(X)$ is the potential energy of the Hamiltonian or Lagrangian. We will clarify this.

[R4] **"the matrix in Eq (5) and (6) might be singular."** Instead of numerically inverting the matrix M, we use the closed form expression for the inverse given on line 195, and values are bounded away from 0 because m and λ are parametrized using a $\mathrm{Softplus}$, so M is never singular. Note that since the constraints $\{\Phi(x)_i = 0\}_{i=1}^{C}$ are assumed to be independent, $D\Phi$ has linearly independent columns, and therefore the matrix $\left[D\Phi M^{-1} D\Phi^\top\right]$ is invertible. The same holds for $\left[D\Psi J D\Psi^\top\right]$ in CHNNs.

Clarity. We appreciate that reviewers found the exposition largely clear. We will include some additional background, and move some material as suggested. We will also revise the summary after line 206, and fix the typo in line 188.