We thank the reviewers for their time and thoughtful feedback. Overall, the reviewers thought the paper was well written (**R1**, **R2**, **R4**) and found the theoretical analysis interesting (**R1**, **R4**). They thought the empirical methodology was comprehensive (**R2**) and thorough (**R1**), and they recognised that Second Order Neural ODEs (SONODEs) have application potential (**R3**). We are very grateful for the suggestions on how to improve this work. A key point is that the paper would benefit from further comparisons (theoretical and empirical) between Augmented Neural ODEs (ANODEs) and SONODEs, and a discussion about the settings in which ANODEs are expected to outperform SONODEs and vice versa (**R2**, **R4**). We agree that this would benefit the work, therefore we address this concern and others below. We refer to lines, figures and pages from the manuscript as (L, Fig., p).

Make further comparisons between ANODEs and SONODEs (**R2**, **R4**) - Given SONODEs are a special case of ANODEs (Eq.4, L65), the reviewers ask when ANODEs might outperform SONODEs (**R2**) and if ANODEs can achieve the same performance with adequate data (**R4**). This depends on the task and the expected underlying dynamics. We believe that for tasks where the trajectory is unimportant, and performance depends only on the end points (such as classification), ANODEs might perform better because they are unconstrained in how they use their capacity. To investigate this, we followed **R4**'s suggestion and included ANODEs in the MNIST experiment (as we did for NODEs and SONODEs in Appendix E.3), augmenting along the channels as is done in Dupont 2019. We found that ANODEs achieve the same accuracy as SONODEs (see figure below), with fewer parameters and only one augmented channel. This is consistent with the result from Dissecting Neural ODEs, where ANODEs had a higher accuracy with five augmented channels and approximately the same number of parameters.

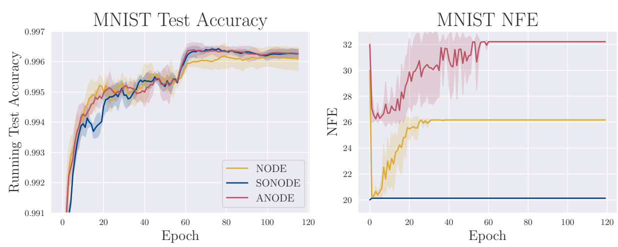

Figure 1: MNIST evaluation: test accuracy and number of function evaluations (NFE).

We expect SONODEs to outperform ANODEs on time-series data when the underlying dynamics is assumed (or known) to be second order (also mentioned by **R1**). In this setting, SONODEs have a unique functional solution and fewer local minima compared to ANODEs. For example, in theory ANODEs can learn the Silverbox task but they are unable to do so (Fig.9). Moreover, better interpretability also makes SONODEs more appropriate for application in the natural sciences, where second order dynamics are common and it is useful to recover the force equation. Additionally, SONODEs train faster as they do not have to learn second order (Fig.6), they are more robust to noise (Fig.7), and will require fewer parameters ($\dot{x} = v$ does not require any parameters). *However*, when the dynamics are not second order we believe ANODEs will perform better as they are not restricted to second order solutions as shown in Fig.8 (the airplane benchmark).

More complex systems, e.g. n-particle dynamics, and higher-order dynamics (**R1**) - Our approach allows for modeling more complex systems, such as n-particle dynamics. For instance, if $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$ and $\mathbf{v} = (\mathbf{v}_1, ..., \mathbf{v}_n)$; then $\dot{\mathbf{x}} = \mathbf{v}$, $\dot{\mathbf{v}} = f^{(a)}(\mathbf{x}, \mathbf{v}, t, \theta_f)$, $\mathbf{v}(t_0) = g(\mathbf{x}(t_0), t_0, \theta_g)$ models the dynamics. However, even simple multi-particle systems are highly sensitive to initial conditions (chaotic) and it becomes computationally intractable to solve the problem to acceptable precision. Moreover, while in this paper we investigate second order dynamics, SONODEs can indeed be extended to higher order to model richer behaviour. We will investigate this by comparing third order to SONODEs and ANODEs on more difficult modelling tasks, such as the airplane task.

Motivation for SONODEs and detailed comparison to previous work (**R3**) - The paper focuses on the theoretical and empirical analysis of second order behaviour in ANODEs. We note that second order dynamics are common in physics (L27-28), however, we will amend the introduction to better contextualise the importance of this work. Moreover, while we discuss most of the relevant related work, we will consider extending this discussion to include a broader overview of related work.

Can SONODEs represent the function used in Dupont et al. 2019? (**R4**) - Yes, we demonstrate SONODEs on the $\mathrm{g}_{1\mathrm{d}}$ and g (p3,p4). However, we use the names compact parity problem (as we consider the generalised parity problem), and nested-n-spheres (name used in Dissecting Neural ODEs). We will add a note on this in the final version.

Confusing that 'a' stands for both acceleration and augmented variable (**R4**) - Whilst we use only 'a' implicitly in the function $f^{(a)}$, we agree with the meta-point that, in dynamics, 'a' often refers to acceleration, which is very relevant in this work. We will amend the text to use a different symbol for the augmented variable to remove this confusion.

Can you explain how Eq 8. is derived? (**R4**) - Start from the state $\mathbf{z} = [\mathbf{x}, \mathbf{a}]$. The velocity can almost be represented by $\mathbf{a}$, but in the original formulation $\mathbf{a}(t_0) = 0$. This is fixed by adding the *constant* $\dot{\mathbf{x}}(t_0)$. Such that $\dot{\mathbf{x}} = \mathbf{a} + \dot{\mathbf{x}}(t_0)$. To get the desired acceleration, $\ddot{\mathbf{x}} = \dot{\mathbf{a}} = f^{(a)}(\mathbf{x}, \mathbf{a} + \dot{\mathbf{x}}(t_0), t, \theta_f)$. Using $\dot{\mathbf{x}}(t_0) = g(\mathbf{x}(t_0), t_0, \theta_g)$ from the SONODE formulation gives Eq 8. This is consistent with the more general expression in Eq 12.