We thank all reviewers for your very useful comments on our paper. Please find our responses to each reviewer below.

To Reviewer #2: Thanks for the thorough reading! Regarding your comments on improvements, please find our response below.

1. We will provide more intuition for high-resolution ODEs in Section 2. At a high level, high-resolution ODEs are finer approximations to NAGs that also consider $O(\sqrt{s})$-order terms in contrast to low-resolution ODEs which only consider $O(1)$-order terms. These finer approximations allow us to investigate more refined aspects of the dynamics.
2. Note that Direct RK admits the convergence rate $O(k^{-2s/(s+1)})$ which can be faster than gradient descent but is slower the optimal $O(k^{-2})$ rate. We will clarify and discuss this explicitly in the final version.
3. Thanks for your suggestion. We will modify our paper accordingly.
4. The gradient correction term is $\frac{k}{k+1} \cdot s\left(\nabla f(x_k) - \nabla f(x_{k-1})\right)$ for NAG-C and $\frac{1-\sqrt{\mu s}}{1+\sqrt{\mu s}} \cdot s\left(\nabla f(x_k) - \nabla f(x_{k-1})\right)$ for NAG-SC. Note the effect of the gradient correction term can be ascertained from High-resolution ODEs but not from low-resolution ODEs. We will add a clarification in this regard in the final version.
5. The two algorithms are the standard NAG-SC and the symplectic discretization of high-resolution ODEs.
6. We will modify it accordingly.
7. We will add more detail and description. Thanks for the suggestion.
8. Note that the coefficient function in Theorem B.1 (Theorem B.2) is a continuous function of s. Therefore the maximum value of this function in Theorem 3.1 (Theorem 3.2) is achieved on the closed interval.
9. Yes, that is a typo. Thanks for pointing out.

To Reviewer #3: Thanks for asking about intuition for symplectic methods. We provide some basic intuition here; we will incorporate these comments in the final version.

One explanation for the superiority on the symplectic method comes from Physics. The symplectic structure is an essential property of Hamiltonian systems, capturing aspects of their geometry that are relevant to invariances of the dynamical flow. For numerical simulation of the Hamiltonian system, the symplectic method involves a forward discretization step and a backward discretization step; together these steps exploit the symplectic geometry such that low-error discretization errors cancel. In contrast, the explicit Euler scheme has two forward discretizations and the implicit Euler two backward discretization; in neither case do the low-order discretization errors cancel. The cancellation of error terms means that larger step sizes can be taken while retaining stability; this is the core of the connection to acceleration.

Consider the linear Hamiltonian system $H(x, y) = \frac{1}{2}(x^2 + y^2)$ whose trajectories are a family of closed curves. After some calculations, one can show that the explicit Euler scheme will diverge to infinity and the implicit Euler scheme will converge to zero. However, the symplectic Euler scheme can guarantee the numerical solution is a closed curve with properly chosen step size.

To Reviewer #4: Thanks for your very positive comments as well as helpful suggestions! We will consider adding numerical experiments in the final version.