We thank the reviewers for their time, thoughtful reviews, and feedback; the suggestions will help us strengthen the paper. (R1) found the paper to be original and insightful and the derivations to be elegant; (R2) agrees that this is a new research direction that deserves to be explored; (R1) and (R3) noted the clarity of the writing. We thank the reviewers for their kind words. The concerns raised by the reviewers focus on how our methods relate to other methods in the literature and the convergence rates. We address each in turn, along with other minor questions.

Relationship to other methods. Firstly, we added *all* suggested references, including related work from Hedy Attouch; we thank the reviewers for bringing these to our attention. We have expanded the discussion of other methods in the literature including Condat-Vu and Hamiltonian perspectives on acceleration. (R3) noted that we might obtain other existing methods as a discretizations of Hamiltonian descent. To address this we have added a derivation of PDHG from the Hamtilonian descent ODE to the appendix; unfortunately Condat-Vu does not fit neatly into our framework (at the moment at least).

(R3) had some concerns about the discretization trick we used to derive a connection between ADMM and the Hamitlonian ODE. One can view it from the other direction—the Hamiltonian ODE can be recovered by taking the step-size in ADMM to zero. The 'trick' we use is standard, see, e.g., the cited work by Wilson, Recht, and Jordan, 2018. Finally, we provide two other 'vanilla' discretizations that do not use that particular trick.

Convergence questions. Regarding convergence rates of ADMM, (R1) asked whether this Hamiltonian perspective on ADMM will yield a distinct convergence rate analysis. This is an interesting question that we must leave to future work. Ultimately, each discretization scheme is individually analyzed and ADMM has been extensively studied in the literature where an $O(1/k)$ rate can be obtained, e.g., by He and Yuan, 2012 (we have added this reference too).

(R3) noted that the $O(1/k)$ convergence rate in our analysis is not matched by experiment. Our analysis is a worst-case analysis that does not assume strong convexity - for some problems it may well do better than the worst-case rate. In particular the two examples we presented have (local) strong convexity - we have added an additional proof to the appendix that shows linear convergence under (global) strong convexity. Now the experimental results and the theory match for the strongly convex case.

(R1) suggested extending the analysis to non-smooth Hamiltonians and pointed out some potential difficulties, along with a reference. We agree that this would be very useful work and would like to tackle it in the future, although the difficulties pointed out are significant.

Additional experiments. (R2) asked for examples on higher dimensional data. We have added additional runs of the existing experiments scaling to higher dimensions - we include here a sneak preview of the results of the same problem as example 1 where the data is 1000×1000 instead of 50×50, in Figure 1.

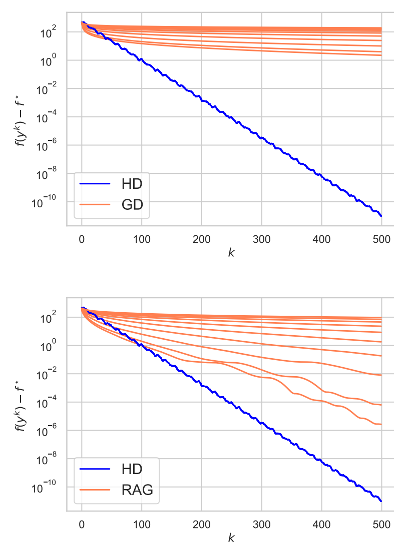

Figure 1: Higher dimensional regularized least squares example. Problem (14) in original draft.

Generality of composite optimization. (R3) was concerned that the composite optimization problem was not sufficiently general to warrant interest. This problem (the sum of two objectives related by a linear mapping) has been studied extensively in the literature and many problems can be written in this form, including cone programs, regularized loss minimization, etc. Since it has so many applications, many algorithms have been developed to solve it, including ADMM, FISTA, PDHG, Condat-Vu etc. We think this is, in fact, one of the most important problems in optimization.

Importance of affine invariance. (R3) did not agree that the affine invariance property was important. Affine invariance provides robustness to poor conditioning without explicit knowledge of the conditioning; it is one of the main advantages that second-order methods have over first-order methods. For this reason, it is interesting to have a first-order method (Hamiltonian descent) that is affine invariant. We refer (R3) to Boyd and Vandenberghe, 2004, Sec 9.5 where the importance of this property is discussed. In fact, the message in Figure 1 of the original draft was to demonstrate numerically the practical impact of affine invariance; we generated a sequence of problems with worsening conditioning and showed that both standard and accelerated gradient techniques suffer greatly when the conditioning worsens, but our technique is unaffected. Since our technique is unaffected by the worsening conditioning it would stand to reason that for some problems it would do better than an accelerated gradient technique.