Reviewer 1 Thank you for the comments on how to improve the notation, and the specified typos. We will incorporate these changes. Our responses to your main questions are below.

- We will restrict our discussion to strongly convex functions rather than strictly convex functions in the differential geometry section, as we don't currently handle the extra complexity from strict convexity in our discussion. Thank you for pointing this out.

- The view of the gradient map as a diffeomorphism is a useful way of visualizing what's going on, we will mention this in the updated draft. We will provide the reference regarding geodesics that you request as well.

- We do apply the Bregman divergence machinery to both ϕ and f. ϕ is used in the introductory sections rather and f throughout because we technically use the conjugate of f, not f its self, in the main change of geometry, and we wished to avoid notation involving conjugate-of-conjugate operations. We will update the wording to make this clearer.

- At line 154, the flatness ensures the solution has a simple form, you are correct that a solution will always exist even without flatness.

Reviewer 2 - The research avenue of understanding acceleration by discretizing continuous time dynamics is certainly one of the leading approaches currently being pursued. We have followed this literature carefully, as and far as we are aware all the existing approaches suffer from an "ad-hoc"-ary problem, where going from the continuous to discrete form requires a non-obvious or non-standard discretization method. The Approximate duality gap technique suffers from this problem, where as they state "... we can introduce an additional gradient step whose role is to cancel out the discretization error by reducing the upper bound.". Although certainly a promising approach, we don't consider this a satisfying explanation for the performance of acceleration, since it would be difficult to arrive at Nesterov's method via this path without already being aware of its functional form (i.e. working backwards). In contrast, our technique just relies on replacing the Bregman divergence used in the proximal point method with another more easily computable divergence, which is a natural step.

"As [1] showed, the modern formulation of estimation sequence has strong power to explain acceleration."' - We don't believe that the mentioned paper refutes our statement about estimate sequences, as they only derive composite-dual averaging via their discretization technique, not the simpler Beck and Teboulle method, which is the explicit claim we make. Likewise they do not apply their technique to variance reduced objectives, which is problematic to address using continuous time dynamics. The accelerated form of Lan that we build upon does apply to variance reduced problems.

Addressing point (2), our technique does lead to a ODE whose forward Euler discretization is naturally a discrete accelerated method, unlike the approach in [1]. We believe the explanation that our method provides, i.e. that Nesterov's method is the proximal point method in disguise, provides substantial insight that other explanations lack.

We hope that Reviewer 2 will revise his judgement based on our comments, although we acknowledge that the utility of new explanations of existing algorithms is subjective.

Reviewer 3 "To me, the weakest aspect of the paper is that the new interpretation of the accelerated gradient method still does not explain why it achieves acceleration" This is certainly the fundamental question. The high-level approach we attempt in this work is to show that it is just a form of the proximal point method. We believe the proximal point method is "intuitively" fast, but of course this is subjective.

- The mentioned Bubeck et. al. paper is perhaps the most unique approach to acceleration developed in recent years. We don't discuss it in detail in our work as the approach of finding points in the intersection of two balls is quite distinct from the proximal approach we look at, and it is not equivalent to Nesterov's method.

- The linear coupling paper by Allen-Zhu and Orecchia provides an interpretation of acceleration as linearly interpolating between a primal and a mirror step. This interpretation uses an equational form of acceleration similar to the one we build upon. It's certainly another valid interpretation, but we believe it lacks explanatory power, as it is unclear *why* such an interpolation would give an accelerated rate. As they state, it requires a substantial amount of analysis to derive the interpolation constant, which is not the case in our version, where the τ factor is directly given by the change-of-geometry.

- The heavy ball section could potentially be cut. We kept it in as we were often receiving questions about the relation to the heavy ball method whenever we discussed our work with others. We will move it to the appendix.

- You are correct that the biorthogonal condition is just a change of variables for the tangent spaces. It's notable just because of it's particular simplicity for these two coordinate systems.

Thank you for the remaining detailed comments. We do not have space here to address them one-by-one but we will update the manuscript to fix all the mentioned issues and provide clarification where requested.