We thank the reviewers for their thoughtful feedback. It seems we have not sufficiently clarified the context and scope of our contributions, so we begin with general comments in this direction. We will update the introduction of our paper accordingly. We thank the reviewers for bringing this to our attention.

Clarifying the context and scope of our contributions. SVGD is markedly different from sampling methods based on diffusions, such as Langevin or HMC, in that it constructs a sequence of *deterministic* mappings whose composition approximately pushes forward an initial distribution to the target distribution. The discovery of its interpretation as a gradient flow of the KL divergence in the Wasserstein space of probability measures, as well as a connection to the rich mathematical theory of interacting particle systems, has led to great interest among researchers. (As noted by **R2**, a gradient flow interpretation lends itself easily to convergence analysis.) However, despite fervent activity, the prevailing perspective has not provided sufficient theoretical understanding for SVGD to overcome its known problems, such as mode collapse (**R3**) or lack of guidance on how to choose an appropriate kernel, and consequently diffusion-based algorithms remain the dominant choice for applications.

In this work, we provide a new and stronger theoretical footing for the development of such deterministic mappings. In particular, it allows us to derive improved convergence guarantees for SVGD (including strong uniform ergodicity, a property which is not standard in the literature; **R2**). Additionally, by introducing a simpler conceptual framework in which the properties of a single integral operator, $\mathcal{K}_\pi$, governs the performance of SVGD, our work is the first to demonstrate that the interplay of the kernel and the target distribution is crucial for designing SVGD-like algorithms. As a proof of concept, our proposed algorithm uses a kernel which is carefully designed based on the target distribution. We believe our study will stimulate further work on the design of sampling algorithms, based on deterministic pushforward mappings, which may eventually see the same widespread application as diffusion-based algorithms.

In fact, we do not advocate for LAWGD as the definitive solution of the sampling problem, but rather as the start of a family of interacting particle systems with an interacting potential that depends strongly and non-trivially on the target distribution and furthermore comes with strong theoretical guarantees.

On the role of numerical PDEs. LAWGD establishes a firm bridge between the fields of sampling and numerical PDEs, whereby the main computational bottleneck of our algorithm is the inversion of a differential operator. Although our naïve implementation of LAWGD is not scalable to high dimension (as rightly pointed out by **R2** and **R4**), the problem of efficiently solving high-dimensional PDEs is precisely the target of intensive research in modern numerical PDEs, culminating in a wide variety of methods ranging from ad-hoc but effective neural network approaches to more principled solvers. We view this bridge as one of our core contributions and we hope that attracting the attention of numerical PDE researchers will yield fruitful collaborations for both fields.

Addressing specific comments.

- **R1** and **R4** write that our experiments are not extensive enough. This is an excellent point and we fully agree that more in-depth experiments are required to evaluate the practicality of LAWGD. As mentioned above, however, our goal is not to establish the supremacy of LAWGD by testing it on a battery of challenging high-dimensional instances, but rather to demonstrate that, unlike SVGD, LAWGD has both strong theoretical guarantees and good numerical performance (recall that for non-trivial kernels there is currently no quantitative convergence analysis of SVGD under verifiable assumptions and that even our simple experiments on the mixture of two Gaussians demonstrate a failure of SVGD). This is an indication that our novel perspective could be the correct one to further advance the state-of-the-art for sampling via deterministic mappings.
- **R2** and **R4** note that we do not provide analysis in discrete time. Although discrete time analysis is common for the more established class of diffusion-based sampling algorithms, the understanding of SVGD-like algorithms is still nascent. Following the trend of recent work such as Duncan et al., we work in continuous time in order to develop conceptual understanding regarding the impact of the choice of kernel.
- **R3** raises a number of interesting questions, and we address some here. Just as we have established SVGD as a kernelized gradient flow of the chi-squared divergence, it would indeed be interesting to consider gradient flows of other functionals, such as f-divergences, and to use kernels or develop other approaches to implement them; we mention these directions in our open questions.

  For LAWGD, the choice of kernel is fully determined by the target distribution, and in principle there is no risk of the kernel being mismatched to the target distribution (as in the case of the observed failure of SVGD with RBF kernels). In practice, however, more numerical experiments are necessary to determine if LAWGD suffers from similar problems as SVGD in high dimension.

  We also note your thought-provoking question about a more direct connection between LAWGD and the Schrödinger operator. While we cannot see an obvious connection in the context of sampling (for which $\mathscr{L}$ is more natural), it is possible that it becomes a more central object in particle methods for PDEs without a limiting distribution.