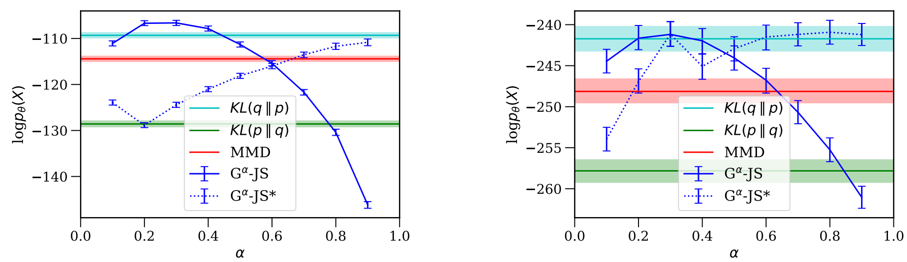
We thank the reviewers for their thoughtful feedback. We are encouraged to see that all reviewers recognise our three contributions and their relation to the NeurIPS audience: introduction of our $\mathrm{JS}^{\mathrm{G}_\alpha}$ variant, justification of interpolation between forward and reverse KL, and application to VAE regularisation with improved empirical results. We are pleased to see that all reviewers found our paper well-written (**R3** has several detailed/useful improvements), well-related to prior work, and easily reproducible. Reviewer consensus also strongly supports the correctness of our paper, with the minor concerns addressed below.

Generative experiments. **R4**, in Fig. 1 we include experiments conducted following the review, estimating model evidence (ME) in comparison to forward KL, reverse KL and MMD (JS being intractable). ME estimates are generated by Monte Carlo estimation of the marginal distribution $p_\theta(X)$ with mean and 95% confidence intervals bootstrapped from 1,000 resamples of estimated batch evidence across 100 *test* set batches. We again emphasise here that we are not looking for SoTA, but *relative* improvement which isolates the impact of the proposed regularisation and extends our analysis of $\mathrm{JS}^{\mathrm{G}_{\alpha'}}$. Hopefully, these steps also address **R2**'s comment on further experimentation.



(a) MNIST (b) FashionMNIST

Figure 1: Log model evidence compared across α values. The increased reconstructive power of $\mathrm{JS}_*^{\mathrm{G}_\alpha}$ does come at a cost to generative performance, however this trend is not consistent in the noisier (b) FashionMNIST dataset. Nevertheless, note that the reconstruction error of $\mathrm{JS}_*^{\mathrm{G}_\alpha}$ for (a) $\alpha > 0.8$ and (b) $\alpha > 0.6$ is still lower than the benchmarks. We also find $0.15 < \alpha < 0.4$ for $\mathrm{JS}^{\mathrm{G}_\alpha}$ is competitive with or better than all alternatives on both datasets.

Correctness. **R3**, L134 should indeed be $\alpha = 0$ for $\mathrm{JS}^{\mathrm{G}_\alpha}$ and $\alpha = 1$ for $\mathrm{JS}_*^{\mathrm{G}_\alpha}$ (see typo mentioned below).

Extension of Nielsen 2019 and related work. W.r.t. **R3**'s question about the Nielsen paper, we reverse the intermediate distribution parameterisation allowing a principled interpolation of forward and reverse KL, we simplify the subsequent closed-form loss to that needed for VAEs, and we demonstrate improved empirical performance against several baselines (application, rather than the theory of Nielsen's papers). We will amend the text to make this clearer. In this regard, we will also extend the related-work discussion, to further highlight the contributions of this work.

Disentanglement Although **R2** and **R4** note the question of disentanglement, we deliberately chose to leave this question for future/separate work in order to keep the message simple—interpolation between forward/reverse KL and the consequences of their associated zero-forcing/avoiding properties on VAE regularisation.

General comments. **R1**, w.r.t. improving reconstruction quality (well-known to be an issue with VAEs) we could see a future line of work using $\mathrm{JS}^{\mathrm{G}_\alpha}$ in nested VAEs (such as Dai and Wipf's *Diagnosing and Enhancing VAE Models*) and recent larger architectures. **R1**, yes we do have a rather important typo in Equation 19 that will be fixed.
R2, using the proposed regularisation has no impact on complexity compared to other VAEs. Equations 23 and 24 can be vectorised and FLOPs are $\mathcal{O}(n)$ where n is the size of the latent space (≤ 32 for our experiments). **R2**, we will make it clearer that hyperparameters for β-VAE and MMD were selected based on the most successful models in the respective papers. W.r.t **R2**'s question regarding Wasserstein Autoencoders, whilst we will consider performing detailed analysis, we expect their performance (at least WAE-MMD) to be comparable to the VAEs regularised with MMD. **R2**, we will add the limiting reconstructions of α from Fig. 4 to the supplementary material as this is one of the avenues we experimented with ahead of submission. **R3**, regarding better explanation of Fig. 1 (in the manuscript) being objectively in favour of $\mathrm{JS}^{\mathrm{G}_\alpha}$, the aim of this figure was to demonstrate that the geometric mixture of Gaussians leads to a larger integral (and therefore regularisation) contribution in areas of low probability for either distribution. The weighting term in forward and reverse KL (and therefore numerically integrated JS) will allow for very little contribution in these regions. **R4**, we will consider moving Figures 3c and 3d to the Supplementary, as space will be used elsewhere. **R3**, **R4**, thank you for the detailed minor comments/suggestions sections, they will be taken into consideration. **R1** and **R4** had suggestions for Fig. 4 which we believe to be valid. Our approach will be to reduce the number of grid entries, therefore expanding subfigure size, and add labelling/caption content to minimise ambiguity.

Broader impact. **R2**, **R3**, we accept that the broader impact statement can be expanded upon and will address this in the redraft. We will emphasise the advantages and disadvantages of using generative models, as well as synthetic (generated) data usage in various data-scarce scenarios and as a mechanism to forge real data.