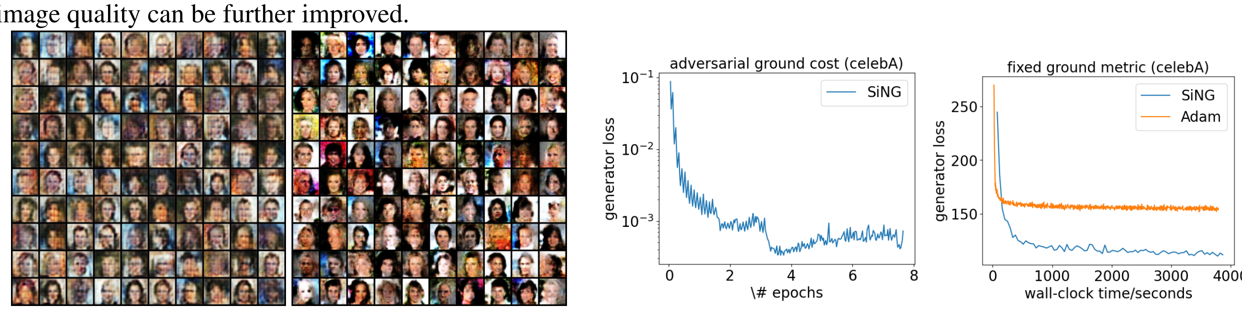
We thank the reviewers for their careful consideration and constructive feedback. Below, please find our responses.

To Reviewer #1. Q1. Computational cost of using Hessian. A1. The reviewer is right: The per-iteration cost of SiNG is larger than SGD-based methods since we use the conjugate gradient method to compute the update direction. However, this additional cost is well justified as the the output of SiNG has a significantly smaller objective value and hence higher solution quality compared to SGD-based methods, when using *the same* wall-clock time (e.g. see figure (iv)). We will add the discussion in our revision. Q2. Image quality compared to [Genevay et al, 2018]. A2. Please see Figure (i) ([Genevay et al, 2018]) and Figure (ii) (our paper). The entropy regularization of the Sinkhorn divergence is set to $\gamma = 100$ as suggested in Table 2 of [Genevay et al, 2018]. The regularization for the constraint is set to $\gamma = 1$ in SiNG. We used ADAM as the optimizer for the discriminators (with step size 10^{-3} and batch size 4000). We can see that the images generated using SiNG are much more vivid than the ones obtained using SGD-based optimizers. We remark that our main goal has been to showcase that SiNG is more efficient in reducing the objective value compared to SGD-based solvers, and hence, we have used a relatively simpler DC-GAN type generator and discriminator (details given in the supplementary materials). If more sophisticated ResNet type generators and discriminators are used, the image quality can be further improved.



(i) Adam-GAN (ii) SiNG-GAN (iii) stability of SiNG (iv) G-loss vs time

Q3. Unbounded cost functions. A3. Thanks for pointing out this interesting work. We will definitely look into it. Q4. Relationship to Sinkhorn EM (sEM). A4. We thank the reviewer for pointing out this interesting work. SiNG and sEM have different problem formulations: In SiNG, the objective is to find a parameterized measure α_θ such that the functional $\mathcal{F}$ (see eq. (1) in our paper) is minimized; On the other hand, sEM considers an entropy regularized optimal transport distance with a parameterized ground cost g_θ (see eq.(3) in their paper). Hence, the Hessian computations involved in SiNG and sEM are quite different. We will discuss this in our revision. Q5. Sinkhorn information matrix (SIM) is defined without debiasing. A5. SIM is defined to be the Hessian of the Sinkhorn divergence in eq. (10). There are two terms in SIM: $\nabla^2_\theta \mathcal{S}(\alpha_\theta, \beta) = \nabla^2_\theta \mathrm{OT}_\gamma(\alpha_\theta, \beta) + \nabla^2_\theta \mathrm{OT}_\gamma(\alpha_\theta, \alpha_\theta)$. We only derived the explicit expression of the first term due to space limitation. The second "debiasing" term is obtained similarly. Both terms do matter and are involved in our computations. We will elaborate on this in our revision. We apologize for the confusion.

To Reviewer #2. Q1. The stability of SiNG. A1. SiNG is stable in both the fixed ground cost case (see figure (iv)) and the adversarial case (see figure (iii)). Q2. How good are the results? A2. Please see our response A2 to reviewer #1 and figures (i) and (ii). Q3. Results for fixed ground cost. A3. We present the comparison of the generator loss (the objective value) vs time plot in figure (iv), where we have used the DC-GAN generator and the squared ℓ_2 distance as ground metric (no encoder). The entropy regularization parameter γ is set to 0.01 for both the objective and the constraint. We can see that SiNG is much more efficient at reducing the objective value than ADAM given the same amount of time. Due to the limitation of space, we do not provide the images here, but we observe that SiNG is producing more diversified images than ADAM (possibly due to mode collapse of ADAM). We will discuss these in our revision.

To Reviewer #3. Q1. SIM for 1-d Gaussian. A1. Please see Theorem 1 of `https://arxiv.org/abs/2006.02572` for the closed form expression of entropy regularized optimal transport between to Gaussians. Using their result, in the simplest case where only the mean of α is parameterized, i.e. $\alpha = \mathcal{N}(\mu(\theta), \sigma^2)$, SIM admits the closed form $\nabla^2_\theta \mathcal{S}(\alpha_\theta, \beta) = 2\nabla^2_\theta \mu(\theta)$. Q2. SIM under ℓ_1 ground cost. A2. This is an interesting question. Right now we are not able to provide SIM in this case: Our results rely on the smoothness of the Sinkhorn potential and hence the smoothness of the ground cost function which does not apply to the ℓ_1 case. Q3. Relations to previous work. A3. We will definitely discuss these important and interesting related works in our revision.

To Reviewer #5. Q1. Wall clock time comparison. A1. The reviewer is correct about this. SiNG has higher per-iteration computational complexity compared to the simple SGD based methods. However, such cost is justified by the efficiency of SiNG in reducing objective value. Please see A1 to Reviewer #1 and figure (iv). Q2. Add confidence interval in the plots A2. We will add confidence in our revision. Q3. Comparison in standard GAN setting. A3. In the standard GAN setting, by discarding the last layer of the discriminator, we can have an encoder of the input image. This encoder will be acting like the adversarial ground cost we trained in eq. (7). We will add comparisons between SiNG and SGD based solvers in standard GAN setting. Thanks. Q4. Details of implementation. A3. We take 20 CG steps to compute the inverse of SIM without explicitly formulating it (only matrix vector product). In our experiments, batch size of 4000 is already sufficient. We will elaborate more on the details of the experiments in our revision.