We have addressed the minor issues pointed out by the reviewers; thank you for these constructive comments and for the encouraging remarks.

In particular, we have corrected some typos as pointed out by R1 and R3. As suggested by R2, we have also replaced the term GAN in some places with "generative models", to emphasize that our work is not directly related to the training of GANs. We wish to leave the reference style intact, as the guidelines allow such a choice. We will also clarify the x-axis units (time in seconds for figure 2, and cpu time of 1 iteration of gradient descent in figure 1).

In addition, it appears to us that R3's score does not match his/her positive review and we wish to ask R3 to consider revising his/her score. More precisely, R3 is mostly concerned with the experiments and empirical evidence presented in the work, although he/she agrees the paper is strong.

We believe that the theory is backed with solid numerical evidence, arguably well beyond what is common in similar papers, please see for example (1).

In particular, the CelebA dataset is much more complex than MNIST or CIFAR. The version we have used consists of RGB 64x64 images, which is 16 times the dimension of MNIST and 4 times the dimension of CIFAR. Moreover the generator network we used here is a Residual Network, considered to be a good baseline. We will consider also adding CIFAR10/100 examples to the final version, as well as a discussion on future research directions.

Please note that in our second experiment with MNIST, which is widely considered an easy dataset, both Gradient descent and Adam performed poorly. This was observed along a wide range of step sizes for each algorithm. This confidently suggests that these two algorithms would also perform poorly on more complex datasets. Our proposed linearized ADMM, on the other hand, consistently shows superior performance in such nonsmooth problem.

Finally, we also would like to highlight the following: As part of the requirements for this edition of NeurIPS, we will provide a polished code. While we can try to be as exhaustive as possible regarding the numerical evidence, we believe that the real impact will be much more than just including one additional dataset in the current paper.

For clarity, we will also add Algorithm 2 in explicit form to the appendix, as the space in the main text is quite limited. We will also add the proof sketch for Theorem 1 to further clarify the role of the near-isometry property of G; in short, when G is near-isometric, the nonconvex problem locally resembles a convex problem.

We hope that we have addressed any concerns about this work and again wish to ask R3 to consider revising his/her score.

References

[1] V. Shah and C. Hegde, "Solving Linear Inverse Problems Using GAN Priors: An Algorithm with Provable Guarantees," *arXiv:1802.08406 [cs, stat]*, Feb. 2018. arXiv: 1802.08406.