Supplementary material for 'RoCGAN: Robust Conditional GAN'

December 15, 2019

1 Introduction

In the following sections we include the supplementary material accompanying the paper 'RoCGAN: Robust Conditional GAN'. The sections are organized as following:

- In sec. 2 we validate our intuition for the RoCGAN constraints through the linear equivalent.
- In sec. 3 an additional experiment is described.

The Fig. 1, 2, 3, 4 include all the outputs of the synthetic experiment of the main paper. As a reminder, the output vector is $[x + 2y + 4, e^x + 1, x + y + 3, x + 2]$ with $x, y \in [-1, 1]$.

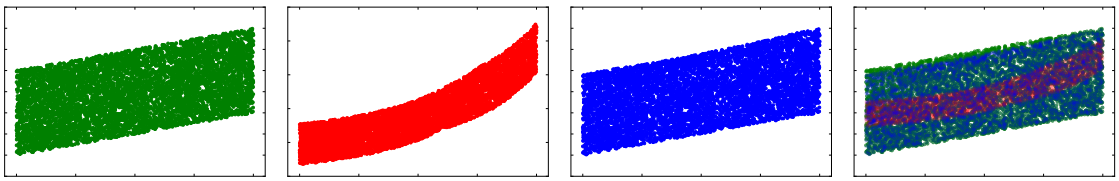

Figure 1: Qualitative results in the synthetic experiment (main paper). Output of the 1^{st} function. From left to right: The target (ground-truth) curve in green, the output of the single pathway network (baseline) in red, the two pathway network in blue and all three overlaid. The output vector of the 1^{st} and the 3^{rd} functions are plotted here with respect to x, the full 3D plot is in the manuscript. All figures in this work are best viewed in color.

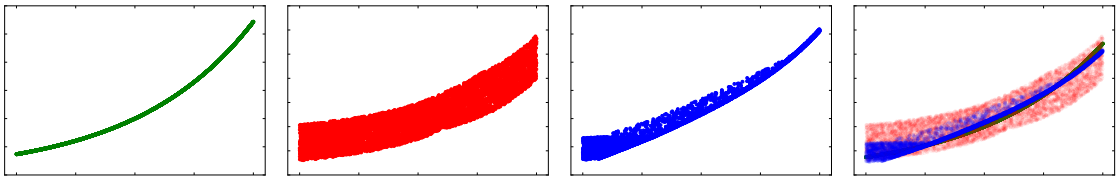

Figure 2: Qualitative results in the synthetic experiment (main paper). Output of the 2^{nd} function. See Fig. 1 for details.

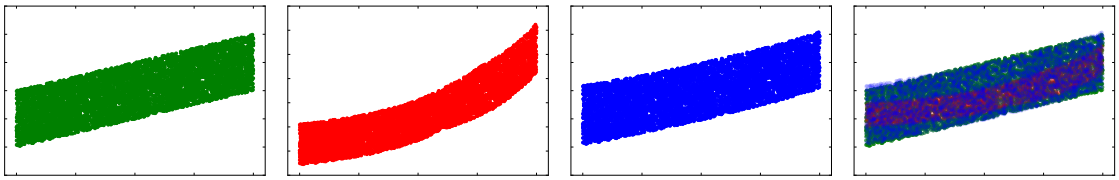

Figure 3: Qualitative results in the synthetic experiment (main paper). Output of the 3^{rd} function. See Fig. 1 for details.

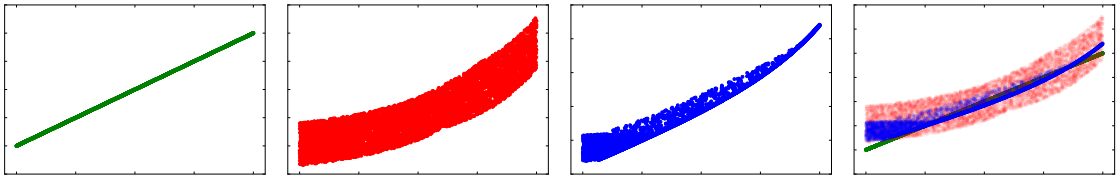

Figure 4: Qualitative results in the synthetic experiment (main paper). Output of the 4^{th} function. See Fig. 1 for details.

2 Linear generator analogy

The exact nature and convergence properties of deep networks remain elusive, however we can study the linear equivalent of deep methods to build on our intuition. To that end, we explore the linear equivalent of our method. Since the discriminator in RoCGAN can remain the same as in the baseline cGAN, we focus in the generators. To perform the analysis on the linear equivalent, we simply drop the piecewise non-linear units in the generators.

We assume a network with two encoding and two decoding layers in this section; all layers include only linear units. The symbols $\boldsymbol{W}_l^{(G)}$ with $l \in [1, 4]$ are the l^{th} layer's parameters (reg pathway). The linear autoencoder (AE) has a similar structure; $\boldsymbol{W}_l^{(AE)}$ denote the respective parameters for the AE. We denote with $\boldsymbol{X}$ the input signal, with $\boldsymbol{Y}$ the target signal and $\hat{\boldsymbol{Y}}$ the AE output, $\tilde{\boldsymbol{Y}}$ the regression output. Then:

$$\hat{\boldsymbol{Y}} = \underbrace{\boldsymbol{W}_4^{(AE)}\boldsymbol{W}_3^{(AE)}}_{\boldsymbol{U}_D^T}\underbrace{\boldsymbol{W}_2^{(AE)}\boldsymbol{W}_1^{(AE)}}_{\boldsymbol{U}_E}\boldsymbol{Y} \tag{1}$$

is the reconstruction of the autoencoder and

$$\tilde{\boldsymbol{Y}} = \underbrace{\boldsymbol{W}_4^{(G)}\boldsymbol{W}_3^{(G)}}_{\boldsymbol{U}_{D,(G)}^T}\underbrace{\boldsymbol{W}_2^{(G)}\boldsymbol{W}_1^{(G)}}_{\boldsymbol{U}_{E,(G)}}\boldsymbol{X} \tag{2}$$

is the regression of the generator (reg pathway). We define the auxiliary $\boldsymbol{U}_{D,(G)}^T = \boldsymbol{W}_4^{(G)}\boldsymbol{W}_3^{(G)}$, $\boldsymbol{U}_{E,(G)} = \boldsymbol{W}_2^{(G)}\boldsymbol{W}_1^{(G)}$, $\boldsymbol{U}_D^T = \boldsymbol{W}_4^{(AE)}\boldsymbol{W}_3^{(AE)}$ and $\boldsymbol{U}_E = \boldsymbol{W}_2^{(AE)}\boldsymbol{W}_1^{(AE)}$. Then Eq. 1 and 2 can be written as:

$$\begin{cases} \tilde{\boldsymbol{Y}} = \boldsymbol{U}_{D,(G)}^T\boldsymbol{U}_{E,(G)}\boldsymbol{X} \\ \hat{\boldsymbol{Y}} = \boldsymbol{U}_D^T\boldsymbol{U}_E\boldsymbol{Y} \end{cases} \tag{3}$$

The AE approximates under mild condition robustly the target manifold of the data [**?**]. If we now define $\boldsymbol{U}_{D,(G)} = \boldsymbol{U}_D$, then the output of the generator $\tilde{\boldsymbol{Y}}$ spans the subspace of $\boldsymbol{U}_D$.

Given that $\boldsymbol{U}_{D,(G)} = \boldsymbol{U}_D$, we constrain the output of the generator to lie in the subspaces learned with the AE.

To illustrate how a projection to a target subspace can contribute to constraining the image, the following visual example is designed. We learn a PCA model using one hundred thousand images from MS-Celeb; we do not apply any pose normalization or alignment (out of the paper's scope). We maintain 90% of the variance. In Fig. 5 we sample a random image from Celeb-A and downscale it[1]; we use bi-linear interpolation to upscale it to the original dimensions. We project and reconstruct both the original and the upsampled versions; note that the output images are similar. This similarity illustrates how the linear projection forces both images to span the same subspace.

3 Experiments

In this section we describe an additional experiment conducted. We follow the same '5layer' network as in the main manuscript; the same training details and dataset. We showcase how our model can be used for a different task than the core experimental results. The task here is to perform sparse inpainting, i.e. the corruption is Bernoulli noise. Specifically, the corrupted data are $(50, 0, 0)$ based on the encoding of the manuscript.

[1]Downsampling is used as a simple corruption. Other corruptions, e.g. gaussian blurring, yield similar results.

(a) original (b) corrupted (c) upscaled (d) reconst-ups (e) reconst-org

Figure 5: Projection to linear subspace. The original image on the left was corrupted (downscaled $\times 4$). The image was upscaled with bi-linear interpolation in (c). Both the original and the upscaled images are projected and reconstructed from a PCA. The reconstruction of the original image ('reconst-org') along with the linearly upscaled ('reconst-ups') image are similar. The simple linear projection demonstrates how constraining the output through a linear subspace can result in a more robust output.

The results for both the faces and the scenes datasets are depicted in table 1. The improvement over cGAN is smaller, however we should note that in this task the difference from the AAE performance is smaller.

Method \ *Experiment*	Faces		Scenes	
	SSIM	*FID*	*SSIM*	*FID*
Baseline-5layer	0.849	47.9	0.692	74.1
Ours-5layer	0.887	31.5	0.697	73.9

Table 1: Quantitative results for the experiment with sparse inpainting (sec. 3).