Spectral Bounding: Strictly Satisfying the 1-Lipschitz Property for Generative Adversarial Networks

Zhihong Zhang[a], Yangbin Zeng[a], Lu Bai[b,*], Yiqun Hu[a], Meihong Wu[a], Shuai Wang[c], Edwin R. Hancock[d]

[a] *Xiamen University, Xiamen, China*
[b] *Central University of Finance and Economics, Beijing, China*
[c] *Beihang University, Beijing, China*
[d] *University of York, York, UK*

Abstract

Imposing the 1-Lipschitz constraint is a problem of key importance in the training of Generative Adversarial Networks (GANs), which has been proved to productively improve stability of GAN training. Although some interesting alternative methods have been proposed to enforce the 1-Lipschitz property, these existing approaches (e.g., weight clipping, gradient penalty (GP), and spectral normalization (SN)) are only partially successful. In this paper, we propose a novel method, which we refer to as spectral bounding (SB) to strictly enforce the 1-Lipschitz constraint. Our method adopts very cost-effective terms of both 1-norm and ∞-norm, and yet allows us to efficiently approximate the upper bound of spectral norms. In this way, our method provide important insights to the relationship between an alternative of strictly satisfying the Lipschitz property and explainable training stability improvements of GAN. Our proposed method thus significantly enhances the stability of GAN training and the quality of generated images. Extensive experiments are conducted, showing that the proposed method outperforms GP and SN on both CIFAR-10 and ILSVRC2015 (ImagetNet) dataset in terms of the standard inception score.

Keywords: Generative Adversarial Networks, 1-Lipschitz constraint, Spectral bounding, Image generation

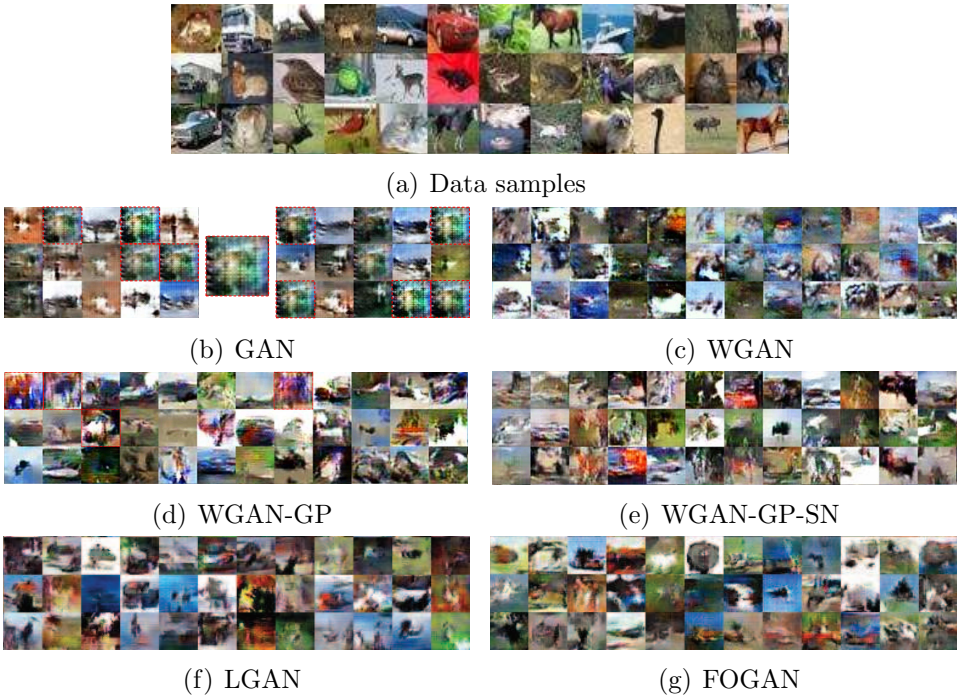
*Corresponding author: Lu Bai
Email address: bailucs@cufe.edu.cn. (Lu Bai)

1. Introduction

One recently emerging technique is the Generative Adversarial Network (GAN) [1], which is widely recognized as one of the most seminal state-of-the-art generative models based on a knowledge of a detailed data distribution. The GAN simultaneously learns a discriminator and a generator by playing a two-player minimax game until a Nash equilibrium is reached. Here the generator can produce a model distribution which is as close to a given real distribution as possible. Meanwhile, the discriminator is taught to distinguish between model and real distributions as accurate as possible, which in turn improves the generator over time in order to be able to "fool" the discriminator. By so doing, at convergence the generative network is capable of generating synthesized data whose distribution can perfectly match the underlying distribution of real data. A family of GAN-based methods has therefore met with considerable success in the kind of data given target distribution.

Many of impressive GAN's variants have these days been proposed for improving the stability of GAN training and avoiding mode collapse, and these alternatives can be roughly categorized into three different classes: a) specially-designed network architectures which replace the multiple fully connected layers in the original GAN [1] with different networks, i.e., deep convolution network and recurrent neural network [2, 3, 4, 5, 6]; b) autoencoder-based GANs which combine variational autoencoders (VAEs) [7] and GAN training [8, 9, 10, 11]; c) alternative objective or distance measure based GANs where the GAN is trained to satisfy different criteria or alternative distance measures between distributions and examples, including Integral Probability Metrics (IPMs) [12] based methods (e.g., Wasserstein GAN (WGAN) [13], Improved Wasserstein GAN via gradient penalty (WGAN-GP) [14] and others [15, 16, 17]), Energy-based Models [18, 19, 20], Least-Square GAN [21], f-GAN [22], BEGAN [23], InfoGAN [24], LS-GAN [25]. Specifically, these improved GANs and their variants have been successfully used in a wide variety of visual tasks, such as image generation [26, 27, 28], and image-to-image translation [29, 30, 31]. However, GAN-based methods are critically dependent on the stability of the training procedure. As a result most fruitful GAN models have to be trained in association with additional stabilization methods [32, 33] (e.g., combining with a specially-designed multi-level network structure [34, 35], or with conditional information [36], etc.).

To provide some visual illustrations, we have evaluated the performance



(a) Data samples

(b) GAN

(c) WGAN

(d) WGAN-GP

(e) WGAN-GP-SN

(f) LGAN

(g) FOGAN

Figure 1: Generated images by (b) GAN, (c) WGAN, (d) WGAN-GP, (e) WGAN-GP-SN, (f) LGAN and (g) FOGAN on CIFAR-10.

of existing GAN-based models on image generation. We experimented with a pure GAN model without any associated stability approaches, and as can be seen in Fig. 1(b), its training fails to converge, leading to low-quality generated images consisting of many repeated patterns. A reasonable explanation for this result is that the KL divergence and Jensen-Shannon divergence used in original GAN cannot work well in measuring the difference between target and model distribution and lack the adequate capacity to avoid vanishing gradients and associated mode collapse in most cases, which has been clearly demonstrated in the literature [13, 15]. In contrast, by incorporating a smooth Wasserstein distance metric and objective, WGAN-based methods perform better than original GAN in terms of quality and diversity of images (see Fig. 1). Nevertheless, the optimization of this objective instructs the discriminator to satisfy the requirement of 1-Lipschitz constraint, which actually bounds the gradient of the discriminator to fall in the range of $[0, 1]$. By using weight clipping in WGAN [13], it is capable of grudgingly abiding by the 1-Lipschitz rule. As widely demonstrated, this approach limits to a large extent

the capability to preserve the discriminative features in the discriminator, which leads to relatively inferior performance (see Fig. 1(c)). Its improved variant, the so-called gradient penalty method (WGAN-GP) [14] imposes an implicit 1-Lipschitz constraint and does not exert a precise bounding, only producing a relatively higher cost for those gradients away from unity. Spectral normalization (WGAN-GP-SN) [37] locks gradients of all matrices (converted from the weights of convolutional layers) to unity. Recently, one interesting attempt of meeting Lipschitz property by manipulating gradients of discriminators as small as possible has been proposed. Specifically, Lipschitz GAN (LGAN) [38] penalizes the maximum of gradients, which is different from WGAN-GP, and First Order GAN (FOGAN) [39] computes a first order information of gradients and is to obtain directly an update direction of gradient decent, similarly encouraging gradients approach to zero. Note that the 1-Lipschitz property, on the other hand, requires gradients to be smaller or equal to unity. As a result of this compromise, these techniques [38, 39] cannot clearly represent detailed texture information (see Figs. 1(d) $\sim$ (g)).

To address the above shortcomings and motivated by the idea, widely demonstrated in the literature [13, 14, 37], of enforcing the 1-Lipschitz condition can significantly improve stability of GAN training, we develop a novel weight rescaling or normalization method referred to as spectral bounding (SB). This enables us to utilize spectral norms to meet the 1-Lipschitz requirement of discriminators. Unfortunately, the calculation of spectral norms is very expensive. We therefore skillfully design a fast and robust algorithm to calculate the upper bound of the spectral norm. Our idea is to bound the spectral norm of the reshaped weight matrix with the product of 1-norm and ∞-norm, which is easily calculated, only requiring $O(n^2)$ computations. By using this approach, we can bound the spectral norm of the discriminator to fall in the interval of [0, 1]. As a consequence, the training procedure condition is constrained to be definitely 1-Lipschitz. We can thereby enhance the stability of the two-player minimax game during GAN training and thus gain an overall improvement in the image generation task when the training process is attempted by balancing the generator and discriminator. In this process, the generator is able to generate samples in an attempt to maximally confuse the discriminator, whereas the discriminator maximizes its classification accuracy until a Nash equilibrium is reached.

In summary, the goal of this study is to explore an ideal explainable GAN-related model which can improve the training stability during reaching the Nash equilibrium from the perspective of imposing the 1-Lipschitz constraint.

The main novelties and contributions can be summarized as follows:

- The proposed spectral bounding method is an original addition to the methodology for WGAN-based models, which strictly meets the requirement of optimizing the Wasserstein-distance metric or measure, namely 1-Lipschitz constraint. Our method differs from other interesting alternatives (e.g., weight clipping, gradient penalty, and spectral normalization, see Table 1 for more details on their differences and comparisons) and proposes a more precise calculation of spectral norms to encourage gradients of discriminative networks to fall in the range of $[0, 1]$.

- We design a cost-effective method for quickly calculating the upper bound of spectral norms, taken the form of the product of 1-norm and ∞-norm. This effectively reduces the computational cost in comparison with a power iteration applied in SN method (which we shall discuss in later sections 4.2.2).

- Extensive experimental results on CIFAR-10 and ImageNet dataset demonstrate that our approach can maintain more successfully the balance between generators and discriminators encountered prior to a Nash equilibrium having been reached. In so doing we can obtain a robust GAN model which accurately captures features of the statistical distribution for data samples used in training.

The remainder of this paper is organized as follows. Section 2 provides an overview of the related literature. Some preliminary material, including a discussion of matrix norms, and explanations of the effectiveness of spectral bounding versus alternative regularization techniques are introduced in Section 3. We present experimental details and results for image generation in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

2.1. Generative Adversarial Networks

Since Goodfellow et al. [1] proposed GANs in 2014, a collection of GAN-based methods have been developed with the specific aim of improving the stability for GAN training. Essentially, GANs attempt to simulate the distribution of a real sample set without supervision. We denote the generator

Table 1: Comparison of realizing 1-Lipschitz property ($\|f\|_{Lip} = \frac{\|f(\mathbf{x})-f(\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} \leq 1$) by different methods.

Method	Implementation	Result
WGAN [13]	Clips weights $\mathbf{W}$ into a range of [-0.01, 0.01]	$\|f\|_{Lip} \ll 1$
WGAN-GP [14]	Adds a gradient penalty $\lambda \cdot \mathbb{E}[\|\nabla f\|_2 - 1]^2$	$\|f\|_{Lip} \approx 1$
LGAN [38]	Adds a max gradient penalty $\lambda \cdot max\|\nabla f\|_2$	$\|f\|_{Lip} \to 0$
SNGAN [37]	Normalizes weights $\mathbf{W}_{sn} = \mathbf{W}/\sigma(\mathbf{W})$	$\|f\|_{Lip} = 1$
FOGAN [39]	Penalizes a first order information of $\|\nabla f\|_2$	$\|f\|_{Lip} \to 0$
Ours	Rescales weights $\widetilde{\mathbf{W}} = \mathbf{W}/\sqrt{\|\mathbf{W}\|_1\|\mathbf{W}\|_\infty}$	$\|f\|_{Lip} \leq 1$

as $G(\alpha)$, where α is the model parameter set. The distributions of generated and real samples are denoted as $\mathbb{P}_\alpha$ and $\mathbb{P}_r$ respectively. A discriminator $D(\beta)$ is employed to determine the differences between the two distributions, $\mathbb{P}_\alpha$ and $\mathbb{P}_r$. This two-player game can be represented by a minimax objective function as follows,

$$\min_{G(\alpha)} \max_{D(\beta)} D(\mathbb{P}_\alpha - \mathbb{P}_r).$$

Here D is a distance measure for the distributions $\mathbb{P}_\alpha$ and $\mathbb{P}_r$. In practice, it has been found that the way in which distance measurement is computed is essential for successful GAN training. Specifically, instead of using the KL-divergence or Jensen-Shannon divergence employed in the original formulation, a smooth Wasserstein distance metric and objective are used, and then the WGAN [13] can stabilize the training process. In other words, it avoids vanishing gradients and mode collapse encountered in the pure GAN. The WGAN also makes the process of balancing the generator and discriminator much easier. As a result it is possible to train the discriminator until optimality is reached, and then gradually improve the generator. Moreover, it provides an advantageous indicator of quality (based on the Wasserstein distance) for the training progress:

$$\begin{aligned} W(\mathbb{P}_\alpha, \mathbb{P}_r) &= \inf_{\gamma \sim (\mathbb{P}_\alpha, \mathbb{P}_r)} \int \|x - y\| d\gamma \\ &= \inf_{\gamma \sim (\mathbb{P}_\alpha, \mathbb{P}_r)} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|] \end{aligned} \tag{1}$$

where γ is the union of the distributions $\mathbb{P}_\alpha$ and $\mathbb{P}_r$. In Eq. (1), it uses the minimum value of the integral value for γ as the distance between the two distributions. Because it is highly intractable for the infimum in Eq. (1), it

can be transformed into a more tractable equation:

$$W(\mathbb{P}_\alpha, \mathbb{P}_r) = \sup_{\|f\|_{Lip} \leq 1} \mathbb{E}_{x \sim \mathbb{P}_r}[f(x)] - \mathbb{E}_{y \sim \mathbb{P}_\alpha}[f(y)]. \tag{2}$$

In Eq. (2), f plays the role as the discriminator and is required to satisfy 1-Lipschitz property.

Algorithm 1 Spectral Normalization [37]

Require:

Parameter $\mathbf{W} = \{\mathbf{W}^l \in \mathbb{R}^{n_l \times m_l}\}_{l=1,2,...,L}$;

Random vector $\tilde{\mathbf{u}}_l \in \mathbb{R}^{d_l}$ for $l = 1, \cdots, L$ sampled from isotropic distribution.

Ensure:

Rescaled $\widehat{\mathbf{W}} = \{\widehat{\mathbf{W}}^l \in \mathbb{R}^{n_l \times m_l}\}_{l=1,2,...,L.}$

1: **for** $l = 1; l \leq L; l++$ **do**

2: Apply power iteration method to a unnormalized parameter matrix $\mathbf{W}^l$:

$\tilde{\mathbf{v}}_l \leftarrow (\mathbf{W}^l)^T \tilde{\mathbf{u}}_l / \|(\mathbf{W}^l)^T \tilde{\mathbf{u}}_l\|_2$

$\tilde{\mathbf{u}}_l \leftarrow \mathbf{W}^l \tilde{\mathbf{v}}_l / \|\mathbf{W}^l \tilde{\mathbf{v}}_l\|_2$

3: Calculate $\bar{\mathbf{W}}^l_{sn}$ with the spectral norm:

$\bar{\mathbf{W}}^l_{sn}(\mathbf{W}^l) = \mathbf{W}^l / \sigma(\mathbf{W}^l)$,

where $\sigma(\mathbf{W}^l) = \tilde{\mathbf{u}}_l^T \mathbf{W}^l \tilde{\mathbf{v}}_l$.

4: Update $\mathbf{W}^l$ with SGD on mini-batch dataset $\mathcal{D}_M$ with a learning rate α:

$\widehat{\mathbf{W}}^l \leftarrow \mathbf{W}^l - \alpha \nabla_{\mathbf{W}^l} l(\bar{\mathbf{W}}^l_{sn}(\mathbf{W}^l), \mathcal{D}_M)$

5: **end for**

To enforce the 1-Lipschitz constraint, several insightful approaches have been adopted (see Table 1). The first of these is weight clipping used in WGAN [13], whereby the weights of the discriminative network are clipped to fall within a pre-specified interval after updating during training. This method is acknowledged as a "simple" and expedient way to realize a Lipschitz constraint, which in some cases lacks the capacity to cope with vanishing gradient and mode collapse [14]. To solve this problem, an additional gradient penalty term is added to the objective function of the discriminator, and this leads to an improved version of WGAN referred to as WGAN-GP [14]. In contrast to weight clipping, WGAN-GP directly requires the gradients to be close to unity by introducing an extra loss term. The resulting WGAN-GP

model encourages the discriminator to be 1-Lipschitz along the path between the generated samples and real samples. However WGAN-GP is an implicit constraint which cannot offer a sufficiently precise bounding in the parameter space. In contrast, LGAN directly penalize the Lipschitz constant which is equivalent to the maximum scale of $\|\nabla f\|_2$ [40], making the resulting LGAN model manipulate the gradients of discriminators approach to zero. Such an explicit idea of penalizing the Lipschitz constant in practice is usually comparable with gradient penalty used in WGAN-GP. Interestingly, FOGAN has proposed a new critic based on first order information of Wasserstein divergence to replace implicit or explicit penalties employed in WGAN-GP or LGAN, which is capable of fulfilling the requirements for unbiased steepest descent updates when updating weights of discriminators. However, we can notice that the penalty of first order information based on $\|\nabla f\|_2$ likewise encourages the norms of gradients of the discriminator close to zero. Together these techniques [38, 39] can to some extent improve the GAN training, but they do not strictly satisfy the 1-Lipschitz property.

SNGAN [37] has thus been proposed as a better way to realize the 1-Lipschitz property for the discriminator. In SNGAN, spectral normalization is employed to stabilize GAN training. To do this SNGAN performs spectral normalization on the weights of the discriminator. The 1-Lipschitz property is imposed by rescaling the parameter value with the max singular value of a reshaped parameter matrix. To do this SNGAN normalizes the spectral norm of the reshaped parameter matrix $\mathbf{W}$ obtained from the weight tensor of the discriminator so as to satisfy the condition $\sigma(\mathbf{W}_{sn}) = 1$:

$$\mathbf{W}_{sn} = \mathbf{W}/\sigma(\mathbf{W}) \tag{3}$$

where $\sigma(\mathbf{W})$ is the largest singular value of $\mathbf{W}$, and equivalently, the spectral norm of $\mathbf{W}$.

To estimate the largest singular value $\sigma(\mathbf{W})$, a power iteration method detailed in Algorithm 1 was employed in SNGAN. However, SNGAN also suffers some shortcomings in terms of its theoretical underpinning, namely a) SN uses a power iteration method to obtain the max singular value of the reshaped parameter matrix which is computationally expensive, b) if there is a multiplicity in the dominant singular values, the power iteration method may fail, and c) SNGAN locks the spectral norm values of all matrices to 1. Unfortunately, this does not satisfy the theoretical requirements of the WGAN [13].

In summary, WGAN, WGAN-GP, and SNGAN propose interesting solutions for stabilizing GAN training, but they still fail to strictly realize the 1-Lipschitz property. Table 1 briefly concludes the ways of enforcing the 1-Lipschitz rule by different alternatives and their corresponding function in the spectral norms of discriminative function f. Neither of the aforementioned methods can therefore be considered as ideal or insightful enough solutions to the optimization of the Wasserstein-distance measurement.

3. Methodology

In this section, we introduce the preliminary concepts and theory of matrix norms, which are a conceptual stepping stone on the path to the theoretical inference of our SB method. Specifically, the goal of SB method is to effectively calculate the product of 1-norm and ∞-norm as the upper bound of spectral norms to realize 1-Lipschitz regulation, which practically requires gradients be smaller or equal to unity. Given a $n \times m$ matrix, $\mathbf{A} \in \mathbb{R}^{n \times m}$, we have

$$\begin{aligned} \|\mathbf{A}\|_2 &= \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\| \\ \|\mathbf{A}\|_1 &= \max_{1 \le j \le m} \sum_{i=1}^{n} |a_{ij}| \\ \|\mathbf{A}\|_\infty &= \max_{1 \le i \le n} \sum_{j=1}^{m} |a_{ij}|. \end{aligned} \tag{4}$$

Here, $\mathbf{x}$ is a specified vector, $\|\mathbf{x}\|_2 = 1$, and a_{ij} is the element with row index i and column index j of the matrix $\mathbf{A}$. Obviously, $\|\mathbf{A}^T\|_1 = \|\mathbf{A}\|_\infty$.

The following theorem and its corollary give the relationship between $\|\mathbf{A}\|_2$, $\|\mathbf{A}\|_1$, and $\|\mathbf{A}\|_\infty$.

Theorem 1.
If $\mathbf{A} \in \mathbb{R}^{n \times m}$, then there exists a unit spectral norm n-vector, $\mathbf{z}$, such that $\mathbf{A}^T\mathbf{A}\mathbf{z} = \mu^2\mathbf{z}$, where $\mu = \|\mathbf{A}\|_2$.

Corollary 1.
If $\mathbf{A} \in \mathbb{R}^{n \times m}$, then $\|\mathbf{A}\|_2 \le \sqrt{\|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty}$.

Proof 1.
If $\mathbf{z} \neq 0$, then $\mathbf{A}^T\mathbf{A}\mathbf{z} = \mu^2\mathbf{z}$ with $\mu = \|\mathbf{A}\|_2$. As a result, we have

$$\mu^2\|\mathbf{z}\|_1 = \|\mathbf{A}^T\mathbf{A}\mathbf{z}\|_1 \le \|\mathbf{A}^T\|_1\|\mathbf{A}\|_1\|\mathbf{z}\|_1 = \|\mathbf{A}\|_\infty\|\mathbf{A}\|_1\|\mathbf{z}\|_1.$$

Our method is based on the well-known WGAN [13]. In this work, we propose a spectral bounding method which is completely orthogonal to the two state of the art GAN training stabilization methods discussed above, i.e., gradient penalty [14] and spectral normalization [37]. Our method can constrain the spectral norm of the reshaped weight matrix to fall in an assigned bounded range of $[0, a]$, where a is the assigned upper bound. In the following analysis we only consider the most widely encountered situation, namely $a = 1$.

Here we detail the SB method with 1-norm and ∞-norm. We begin by formulating the discriminator:

$$f(\mathbf{x}) = \sigma(\mathbf{W}_L ... \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_L), \tag{5}$$

where σ is a nonlinear element-wise function which in practice can be either a sigmoid function or a positive threshold (step) function. Additionally, $\mathbf{W}$ is the weight matrix for every layer whose shape is ($input\_channel \times k \times k$, $output\_channel$), where k is the kernel size. $\mathbf{b}$ is the bias term and $\mathbf{x}$ is an input image. In the WGAN, f is expected to be constrained with 1-Lipschitz continuity. For simplicity[1], we can omit σ and $\mathbf{b}$:

$$f(\mathbf{x}) = (\mathbf{W}_L ... (\mathbf{W}_2 (\mathbf{W}_1 \mathbf{x}))) = (\mathbf{W}_L ... \mathbf{W}_2 \mathbf{W}_1)\mathbf{x}. \tag{6}$$

In practice, the spectral norm is usually used to realize the 1-Lipschitz property, whose mathematical definition for function f is:

For arbitrary x and y in the domain, f should be satisfied with,

$$\frac{\|f(\mathbf{x}) - f(\mathbf{y})\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \leq 1. \tag{7}$$

and then we have,

$$\begin{aligned} \frac{\|f(\mathbf{x}) - f(\mathbf{y})\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} &= \frac{\|(\mathbf{W}_L ... \mathbf{W}_2 \mathbf{W}_1)\mathbf{x} - (\mathbf{W}_L ... \mathbf{W}_2 \mathbf{W}_1)\mathbf{y}\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \\ &= \frac{\|(\mathbf{W}_L ... \mathbf{W}_2 \mathbf{W}_1)(\mathbf{x} - \mathbf{y})\|_2}{\|\mathbf{x} - \mathbf{y}\|_2}. \end{aligned} \tag{8}$$

[1]Actually, f is a feedforward neural network made up of affine transformations and nonlinear pointwise functions (e.g., sigmoid, tanh, ReLU and leaky ReLU), which completely satisfy the 1-Lipschitz property, and the bias term in general does not contribute to the gradient of discriminators while back-propagation.

With the basic inequality of norms,

$$\frac{\|(\mathbf{W}_L...\mathbf{W}_2\mathbf{W}_1)(\mathbf{x}-\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} \leq \|(\mathbf{W}_L...\mathbf{W}_2\mathbf{W}_1)\|_2 \frac{\|\mathbf{x}-\mathbf{y}\|_2}{\|\mathbf{x}-\mathbf{y}\|_2},$$
$$\frac{\|(\mathbf{W}_L...\mathbf{W}_2\mathbf{W}_1)(\mathbf{x}-\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} \leq \|(\mathbf{W}_L...\mathbf{W}_2\mathbf{W}_1)\|_2.$$

Furthermore with Eq. 8,

$$\|\mathbf{W}_L...\mathbf{W}_2\mathbf{W}_1\|_2 \leq \|\mathbf{W}_L\|_2...\|\mathbf{W}_2\|_2\|\mathbf{W}_1\|_2,$$
$$\frac{\|f(\mathbf{x})-f(\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} \leq \|\mathbf{W}_L\|_2...\|\mathbf{W}_2\|_2\|\mathbf{W}_1\|_2 \leq 1.$$

Algorithm 2 Spectral Bounding

Require:
Parameter $\mathbf{W} = \{\mathbf{W}_l \in \mathbb{R}^{n_l \times m_l}\}_{l=1,2,...,L}$.
Ensure:
Rescaled $\widetilde{\mathbf{W}} = \{\widetilde{\mathbf{W}}_l \in \mathbb{R}^{n_l \times m_l}\}_{l=1,2,...,L}$
for $l = 1; l \leq L; l++$ **do**
  $\mu_l = \max(abs(\mathbf{W})_l \mathbf{1}_\mu)$ and $\nu_l = \max(\mathbf{1}_\nu{}^T abs(\mathbf{W}_l))$,
  where $\mu_l, \nu_l \in \mathbb{R}$, $\mathbf{1}_\mu \in \mathbb{R}^{n_l \times 1}$, $\mathbf{1}_\nu{}^T \in \mathbb{R}^{m_l \times 1}$
  if $\sqrt{\mu_l * \nu_l} > 1$ **then**
    $\widetilde{\mathbf{W}}_l = \mathbf{W}_l / \sqrt{\mu_l * \nu_l}$
  else
    $\widetilde{\mathbf{W}}_l = \mathbf{W}_l$
  end if
end for

So if we can bound $\|\mathbf{W}_l\|_2$ then we can also indirectly bound $\|f\|_2$. However, the computation of the spectral norm is very expensive, and this motivates us to investigate an efficient solution. Having developed the corollary 1, we can bound the spectral norm of the reshaped weight matrix with the 1-norm and ∞-norm. This offers the advantage that the 1-norm and ∞-norm are both easily calculated, requiring only $O(n^2)$ computations. In practice, we only need to rescale the matrix with the product of 1-norm and ∞-norm to bound the spectral norm value. For instance, for an unregularized and non-reshaped weight matrix $W_l \in \mathbb{R}^{d_{out} \times d_{in} \times h \times w}$, we begin by regarding the

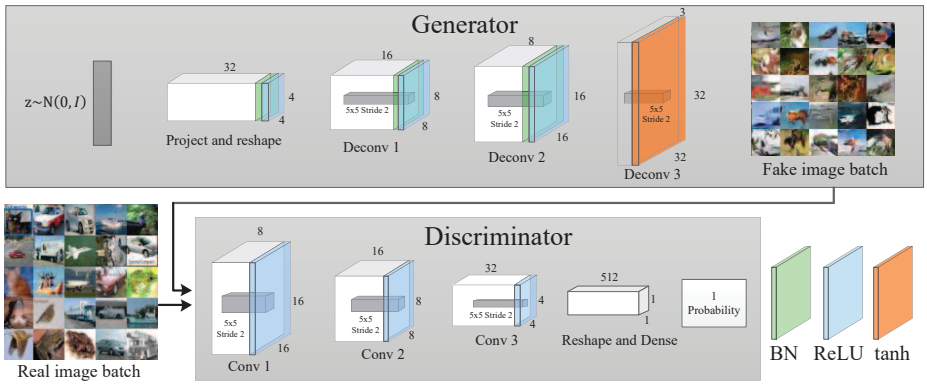


Figure 2: The framework for image generation based on our method.

matrix as a 2-D matrix of size $d_{out} \times (d_{in}hw)$, represented as $W_l \in \mathbb{R}^{n_l \times m_l}$, and then calculate the 1-norm and ∞-norm of itself. If the resulting upper bound of the spectral norm of W_l, i.e., $\sqrt{\|W_l\|_1 \|W_l\|_\infty}$, is greater than unity, we rescale W_l by dividing each element of itself by the corresponding upper bounding, which can enforce the spectral norm to a range of $[0, 1]$. At each iteration during training, the proposed SB method is employed with the set of convolutional weights, obtained before updating the weights of the discriminator. More details can be found in Algorithm 2. Note that the computation is cheap enough in comparison to the calculation of the power iteration step in spectral normalization [37]. Specifically, for one rescaling operation, SNGAN requires at least triple matrix-vector multiplications. In comparison, our method only performs one matrix-vector multiplication for each rescaling operation, which is highly cheaper than SNGAN in computational complexity. More details on computational complexity can be found in Subsection 4.2.2. Additionally, the power iteration used in SNGAN may fail when encountering with multiple roots, and our method definitely sidesteps this bottleneck.

4. Experiments

In this section we detail the experimental configurations and evaluation results. To evaluate the efficiency of our proposed SB method, we conducted a group of unsupervised experiments for generating images on CIFAR-10

[2] and ILSVRC2015 (ImageNet) [3] dataset. We combine WGAN-GP with spectral bounding which we referred to as WGAN-GP-SB and compare its performance against two state-of-the-art methods, namely WGAN-GP [14] and WGAN-GP-SN [37]. This section is organized as follows: we will first present the experimental settings and network architectures, and then report the empirical results and analysis from the perspective of norms' distribution, computational complexity and standard inception score.

4.1. Settings

4.1.1. Database

Following the baselines, we first carried on experiments on the CIFAR-10 dataset to assess the effectiveness of our proposed method. CIFAR 10 consists of 60K 32x32 color images belonging to 10 classes, with 6K images per class. Of all images, we selected 50k ones for training. To make sense of maintaining the efficacy of our method on a large dataset, we also experimented with ImageNet dataset in which the training set includes 1000 classes and approximate 1300 downsampled images of a 32x32 pixel size per class. Therefore, about 1.3 million images are chosen for the training set when trained on ImageNet.

4.1.2. Network Architecture

The framework of our proposed method for image generation is illustrated in Fig. 2. We commence by projecting a 128-dimension Gaussian noise, $z \sim N(0, I)$, into a small spatial convolutional representation processed by a dense network, which in general can be used as the start of the convolution stack. Next, a series of three deconvolution layers, preserving high-level features, synthesize an image of 32x32x3 size. In the discriminator, we use four convolutional layers and a fully connected layer to obtain a probability map from the set of target samples. More details can be found in Table 2. For each of the GAN models studied, we actually employed the standard DCGAN network architecture, which was also applied in [14] and [37], for both the generator and discriminator. Considering the limitation of computing capability of GPU employed in our experiments, the DCGAN network architecture used in our experiments was a toy version in which the channel dimension of both the generator and discriminator is 8 (that of the DCGAN network

[2] `http://www.cs.toronto.edu/~kriz/cifar.html`
[3] `http://image-net.org/small/download.php`

Table 2: Details of standard DCGAN architectures of generator and discriminator for image generation on CIFAR-10 and ImageNet dataset. The $ci(i = 1, 2, 3)$ marked convs are regularized with corresponding methods

(a) Standard Generator

Input: $z \in \mathbb{R}^{128} \sim N(0, I)$. Output: RGB image $x \in \mathbb{R}^{32\times32\times3}$.				
Dense	$4 \times 4 \times 32$	Reshape	BN	ReLU
Deconv	$5 \times 5 \times 16$	Stride=2	BN	ReLU
Deconv	$5 \times 5 \times 8$	Stride=2	BN	ReLU
Deconv	$5 \times 5 \times 3$	Stride=2	tanh	

(b) Standard Discriminator

Input: RGB image $x \in \mathbb{R}^{32\times32\times3}$. Output: 1 probability.			
Conv($c3$)	$5 \times 5 \times 8$	Stride=2	ReLU
Conv($c2$)	$5 \times 5 \times 16$	Stride=2	ReLU
Conv($c1$)	$5 \times 5 \times 32$	Stride=2	ReLU
Reshape	Dense $512 \rightarrow 1$		

architecture is 64). We also experimented with the original WGAN-GP loss function for all models studied, defined as follows:

$$\min_G \max_D \quad \mathbb{E}_{y\sim\mathbb{P}_y}[D(y)] - \mathbb{E}_{x\sim\mathbb{P}_x}[D(G(x))] + \lambda\mathbb{E}_{\hat{x}\sim p_{\hat{x}}}[(\|\nabla_{\hat{x}}D(\hat{x})\|_2 - 1)^2], \tag{9}$$

where λ is a regularization parameter, $\lambda > 0$. $\hat{x}$ is linear recombined sample formed from the real-world sample y and the generated sample $\tilde{x}$. $\hat{x} := \epsilon y + (1 - \epsilon)\tilde{x}$, $\epsilon \sim U[0, 1]$, $\tilde{x} := G(x), x \sim \mathbb{P}_x$. The quantity $\nabla_{\hat{x}}D(\hat{x})$ is the gradient of D for $\hat{x}$. The regularization term, $\mathbb{E}_{\hat{x}\sim p_{\hat{x}}}[(\|\nabla_{\hat{x}}D(\hat{x})\|_2 - 1)^2]$, is used to stabilize the gradient of D close to unity.

4.1.3. Implementation details

To ensure that the superiority of the proposed method is not determined by our specific settings, we experimented with the same optimizer and hyper-parameter settings for all models on CIFAR-10 and ImageNet

dataset. Specifically, we employed the Adam update method as our optimizer in which $\beta_1 = 0.5, \beta_2 = 0.9$. Here β_1 is the exponential decay rate for the 1st moment estimates, and β_2 is the exponential decay rate for the 2nd moment estimates. We set the learning rate and batch size as $2e-4$ and 64, respectively. The critical number for the discriminator was set as 5 and the number of training iterations was set as 100K. There is only one difference of setting hyper-parameter λ between baselines and our model. The λ for the all gradient penalty is 10 when trained by WGAN-GP [14], WGAN-GP-SN [37], LGAN [38] and FOGAN [39]. For our proposed spectral bounding method WGAN-GP-SB, the λ was set to 1 to avoid contradiction between the GP and SB methods, and we relaxed the upper bound to 2 for more freedom and flexibility in the spectral bounding phase. This is primarily because that GP penalizes the gradient term away from unity and a relatively greater λ may severely constrain the gradient in a small enough neighborhood of unity, while our SB method encourages the gradient in a pre-specified range of $[0, a]$. We finally implemented baselines with open source code [4].

4.2. Results

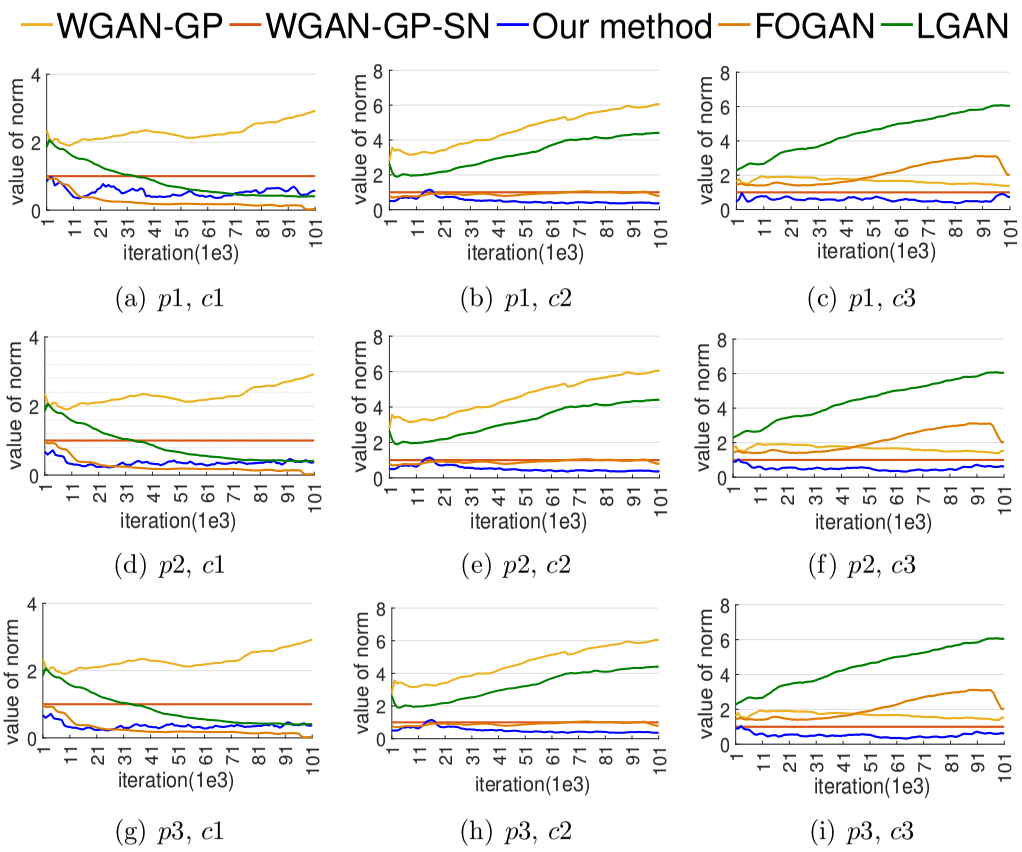
4.2.1. Distribution of Norms

To show the effectiveness of the proposed spectral bounding method, we first visualize the distributions of the spectral norms during the training procedure on CIFAR-10 dataset. The ci $(i = 1, 2, 3)$ are convolution layers which are explained in Table 2. For each convolution layer studied, we record the value of its norms in three phases:

- Phase1 ($p1$): before both spectral bounding and SGD updates.
- Phase2 ($p2$): after spectral bounding but before SGD updates.
- Phase3 ($p3$): after both spectral bounding and SGD updates.

We recorded the change of different norms at an interval of 1K iterations. In Fig. 3, it worth mentioning that the norm values of WGAN-GP were greater than 1 and the norm of WGAN-GP-SN was virtually clamped at 1. Both LGAN and FOGAN force the norms of gradients approach to zero theoretically, but the results indicate that they do not meet the Lipschitz property in all convolutional layers. In contrast, the spectral norm of our SB method was

[4] `https://github.com/igul222/improved_wgan_training.git`



(a) $p1$, $c1$ (b) $p1$, $c2$ (c) $p1$, $c3$

(d) $p2$, $c1$ (e) $p2$, $c2$ (f) $p2$, $c3$

(g) $p3$, $c1$ (h) $p3$, $c2$ (i) $p3$, $c3$

Figure 3: Spectral norm distribution: (a) $p1$, $c1$, (b) $p1$, $c2$, (c) $p1$, $c3$, (d) $p2$, $c1$, (e) $p2$, $c2$, (f) $p2$, $c3$, (g) $p3$, $c1$, (h) $p3$, $c2$, (i) $p3$, $c3$ ($p1$, $p2$ and $p3$ are the three phases defined in Section 4.2.1, $c1$, $c2$ and $c3$ are the convolution layers introduced in Table 2).

completely restricted to the range of $[0, 1]$ for all convolutional weigh matrices in all phases, strictly meeting the 1-Lipschitz property. Notably, the 1-Lipschitz property requires the spectral norms of gradients of discriminators to be smaller or equal to unity. Thus, these results, while preliminary, suggest that our SB method not only is consistent with the related mathematically reasoning, but also can strictly satisfy the requirement of optimization of Wasserstein-distance metric.

4.2.2. Computational Complexity

We next compared the training time of all baselines with exception of GAN and WGAN, and further analyze the relationship between the corresponding

Table 3: Average inception scores for 5 trials on CIFAR-10 and ImageNet dataset.

Methods	CIFAR-10		ImageNet	
	Mean	Std	Mean	Std
GAN [1]	2.7719	0.2051	2.6021	0.1865
WGAN [13]	3.2928	0.2411	2.8655	0.2215
WGAN-GP [14]	3.7446	0.0500	3.6958	0.2142
WGAN-GP-SN [37]	3.8834	0.1383	3.7752	0.2174
LGAN [38]	3.7822	0.1334	3.7569	0.1568
FOGAN [39]	3.9271	0.2266	3.8122	0.2855
Our method	**4.0029**	0.1663	**3.8734**	0.2307

results and theoretical aspects. Fig. 4 shows the average quantitative computational time per update by different methods. Note that WGAN-GP and its variant LGAN slightly faster than other methods. However, our spectral bounding also performs in a relatively economical manner in comparison with spectral normalization. The evidence underpinning this observation consists of two categories of reasons. According to Algorithm 2, our SB method only requires $n(m-1)+(n-1)m$ addition operations to obtain μ and ν. For spectral normalization as shown in Algorithm 1, $2nm+2n+2m$ multiplications, $2nm$ divisions and $2nm-2$ additions are required to obtain $\mathbf{u}$ and $\mathbf{v}$ (about three times computational complexity of our method). Moreover, additional computations were required for computing $\sigma(\mathbf{W})$, the largest singular value of a convolution weight matrix, with power iteration employed in the spectral normalization. On the other hand, the computation of FOGAN requires expensive computational resources to carry out the first order derivative of gradients of discriminators. Therefore, the proposed SB method is far less computationally expensive than both SN and FOGAN.

4.2.3. Inception Scores

We now report some quantitative results for all models and employ *inception score* as a measure of assessing the quality of the generated images by these alternatives. The *inception score* was introduced by Salimans *et al* [2] and is defined as follows:

$$I(\{X_n\}_{n=1}^N) := exp(\mathbb{E}[D_{KL}[p(y|x)]||p(y)]). \tag{10}$$

where $p(y) \approx \frac{1}{N}\sum_{n=1}^N p(y|x_n)$ and $p(y|x)$ is a trained convolutional neural network. The higher the inception score, the better the performance of GAN

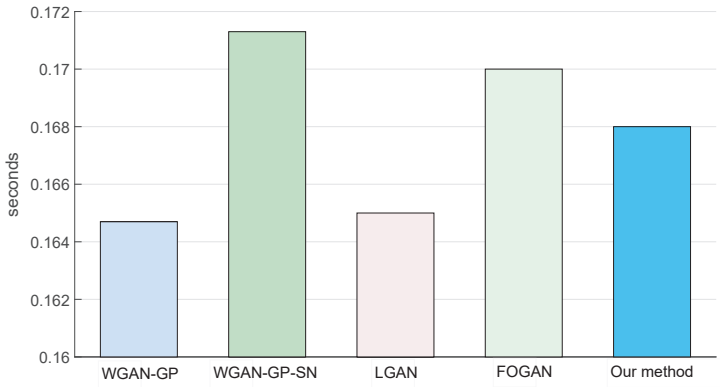


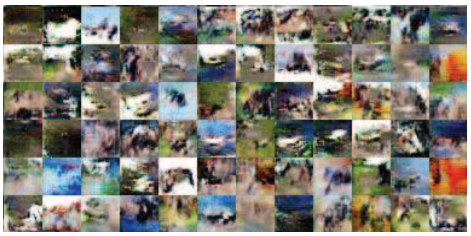
Figure 4: Average computational time for 100k updates.

training on image generation.

We present the generated images obtained using different method on CIFAR-10 and ImageNet dataset in Fig. 5 and Fig. 6, respectively. It should be pointed out that the DCGAN network architecture employed in all models is a toy version with the 8-dimension channel, and therefore may be somewhat difficult to result in a higher level of visual quality. Fig. 7 shows the learning process of standard inception scores with iterations on the two benchmark dataset above produced by different models. It is obvious in Fig. 7 (a) that the learning curve of other methods plateaued around 50Kth iteration, while our method kept increasing even afterward. The similar observation can also be found in Fig. 7 (b). This result may be explained by the fact that we employed $\lambda = 1$ for gradient penalty in our proposed method rather than $\lambda = 10$ used in other baselines, and such a smaller penalty indicates that the performance of our model is inferior to that of other models in the beginning of the training phase. The main purpose of this setting is to avoid contradiction between the GP and our SB method and fairly access the performance of our method. However, the optimal performance of our method in the rest of training demonstrates the effectiveness of our proposed spectral bounding, strictly satisfying the Lipschitz property. Table 3 represents the average inception scores for 5 trials on the two benchmark dataset in which our proposed spectral bounding method WGAN-GP-SB performs better than other models in both CIFAR-10 and ImageNet datasets. In addition, our method is comparable to the alternatives when gauged by the standard deviation indicating the robustness of a model studied. Taken together these quantitative and qualitative results verify the effectiveness of our spectral

(a) Our method

(b) GAN



(c) WGAN

(d) WGAN-GP

(e) WGAN-GP-SN

(f) LGAN

(g) FOGAN

Figure 5: Generated images by different methods on CIFAR-10.

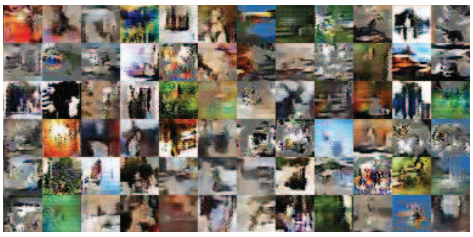
bounding method.

5. Conclusions

(a) Our method

(b) GAN

(c) WGAN



(d) WGAN-GP

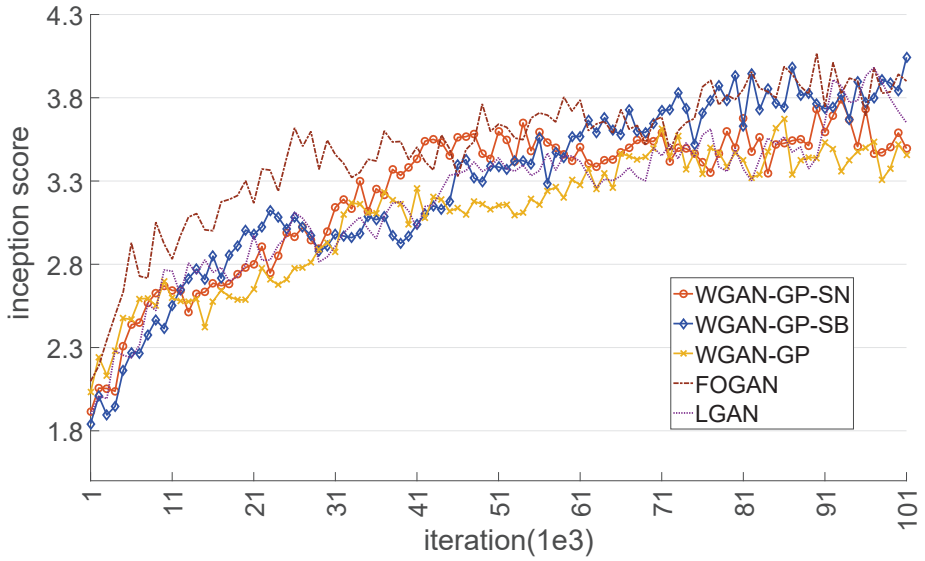
(e) WGAN-GP-SN

(f) LGAN

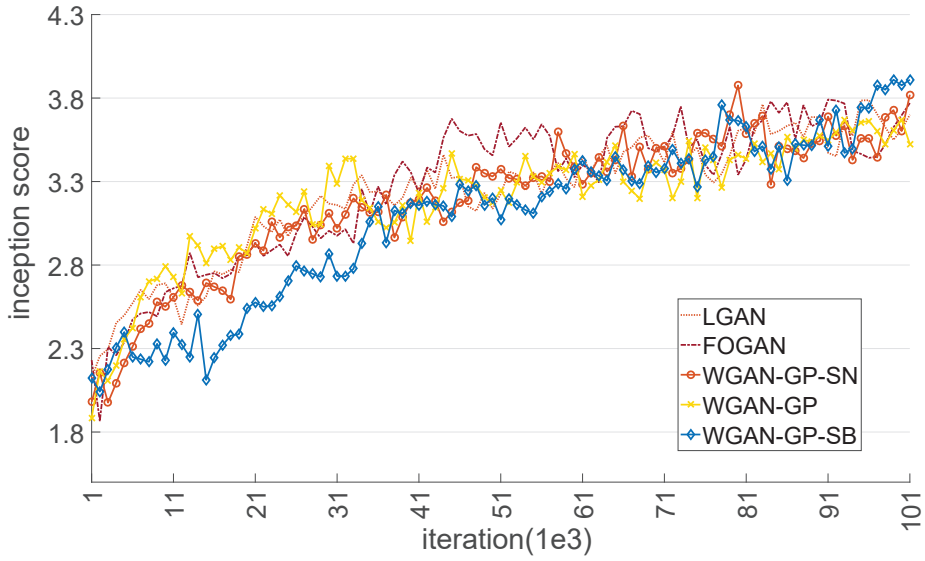
(g) FOGAN

Figure 6: Generated images by different methods on ImageNet.

In this paper, we first have provided the theoretical explanations between several insightful attempts of enforcing 1-Lipschitz continuity used in the literature and the stability degree of GAN's trainings. Inspired by the idea



(a) CIFAR-10.



(b) ImageNet.

Figure 7: Learning curves for unsupervised image generation in terms of inception score for WGAN-GP, WGAN-GP-SN, LGAN, FOGAN and WGAN-GP-SB on CIFAR-10 and ImageNet.

of understanding the stability improvements of deep neural networks optimization with the Lipschitz constraint, we then proposed a spectral bounding

method which can stably constrain GAN training and strictly meet the 1-Lipschitz property. To efficiently calculate the upper bound of the spectral norm, we have presented an effective method by bounding the spectral norm of the reshaped convolutional weight tensor by combining information conveyed by the 1-norm and ∞-norm. This reduces the computational cost and effectively improve the stability of GAN training. Experimental results clearly demonstrate the superiority of our proposed method for generating images and verify the consistency between mathematical reasoning and practical performance. Specifically, the results of training time demonstrate that the proposed method requires less computation than spectral normalization. In addition, our method performs better than several state-of-the-art methods when measured in terms of the quality of synthesized images as gauged by the inception score and together with the subjective visual content. Taken together, the study has gone some way towards enhancing the understanding of Lipschitz constraint and offered some significant insights into the realization of this constraint, and the superiority of which has also been evaluated theoretically and experimentally. Further research should be done to experiment with our method on more large-scale and complex datasets.

Acknowledgment

This work is supported by the Research Funds of State Grid Shaanxi Electric Power Company and State Grid Shaanxi Information and Telecommunication Company (contract no.SGSNXT00GCJS1900134), and the Health joint fund of the Provincial Department of Science and Technology(No.2015J01534).

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. ngio, Generative adversarial nets, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2014, pp. 2672–2680.

[2] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2016, pp. 2234–2242.

[3] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in: ICLR, 2016.

[4] L. Zhou, B. Xiao, X. Liu, J. Zhou, E. R. Hancock, et al., Latent distribution preserving deep subspace clustering, in: 28th International Joint Conference on Artificial Intelligence, York, 2019.

[5] X. Bai, H. Yang, J. Zhou, P. Ren, J. Cheng, Data-dependent hashing based on p-stable distribution, IEEE Transactions on Image Processing 23 (12) (2014) 5033–5046.

[6] K. Gregor, I. Danihelka, A. Graves, D. Rezende, D. Wierstra, Draw: A recurrent neural network for image generation, in: International Conference on Machine Learning (ICML), 2015, pp. 1462–1471.

[7] D. P. Kingma, M. Welling, Auto-encoding variational bayes, in: 2nd International Conference on Learning Representations (ICLR), 2014.

[8] D. P. Kingma, S. Mohamed, D. J. Rezende, M. Welling, Semi-supervised learning with deep generative models, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2014, pp. 3581–3589.

[9] L. Zhou, X. Bai, X. Liu, J. Zhou, E. R. Hancock, Learning binary code for fast nearest subspace search, Pattern Recognition 98 (2020) 107040.

[10] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2223–2232.

[11] O. Bousquet, S. Gelly, I. Tolstikhin, C.-J. Simon-Gabriel, B. Schoelkopf, From optimal transport to generative modeling: the vegan cookbook, arXiv preprint arXiv:1705.07642.

[12] A. Müller, Integral probability metrics and their generating classes of functions, Advances in Applied Probability 29 (2) (1997) 429–443.

[13] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: International Conference on Machine Learning (ICML), 2017, pp. 214–223.

[14] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville, Improved training of wasserstein gans, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2017, pp. 5769–5779.

[15] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, in: ICLR, 2017.

[16] X. Bai, J. Zhou, A. Robles-Kelly, Pattern recognition for high performance imaging, Pattern recognition 82 (2018) 38–39.

[17] X. Bai, C. Yan, H. Yang, L. Bai, J. Zhou, E. R. Hancock, Adaptive hash retrieval with kernel based similarity, Pattern Recognition 75 (2018) 136–148.

[18] Z. Dai, A. Almahairi, P. Bachman, E. Hovy, A. Courville, Calibrating energy-based generative adversarial networks, arXiv preprint arXiv:1702.01691.

[19] T. Kim, Y. Bengio, Deep directed generative models with energy-based probability estimation, arXiv preprint arXiv:1606.03439.

[20] J. Zhao, M. Mathieu, Y. LeCun, Energy-based generative adversarial network, arXiv preprint arXiv:1609.03126.

[21] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, S. P. Smolley, Least squares generative adversarial networks, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), IEEE, 2017, pp. 2813–2821.

[22] S. Nowozin, B. Cseke, R. Tomioka, f-gan: Training generative neural samplers using variational divergence minimization, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2016, pp. 271–279.

[23] D. Berthelot, T. Schumm, L. Metz, Began: Boundary equilibrium generative adversarial networks, arXiv preprint arXiv:1703.10717.

[24] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2016, pp. 2172–2180.

[25] G.-J. Qi, Loss-sensitive generative adversarial networks on lipschitz densities, arXiv preprint arXiv:1701.06264.

[26] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, Cvae-gan: fine-grained image generation through asymmetric training, CoRR, abs/1703.10155 5.

[27] X. Bai, H. Zhang, J. Zhou, Vhr object detection based on structural feature extraction and query expansion, IEEE Transactions on Geoscience and Remote Sensing 52 (10) (2014) 6508–6520.

[28] H. Zhang, X. Bai, J. Zhou, J. Cheng, H. Zhao, Object detection via structural feature selection and shape model, IEEE transactions on image processing 22 (12) (2013) 4984–4995.

[29] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, arXiv preprint.

[30] C. Wang, X. Bai, S. Wang, J. Zhou, P. Ren, Multiscale visual attention networks for object detection in vhr remote sensing images, IEEE Geoscience and Remote Sensing Letters 16 (2) (2019) 310–314.

[31] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, E. Shechtman, Toward multimodal image-to-image translation, in: Proc. Adv. Neural Inf. Proc. Syst. (NIPS), 2017, pp. 465–476.

[32] A. Nguyen, J. Yosinski, Y. Bengio, A. Dosovitskiy, J. Clune, Plug & play generative networks: Conditional iterative generation of images in latent space, arXiv preprint arXiv:1612.00005.

[33] B. Xiao, E. R. Hancock, R. C. Wilson, Graph characteristics from the heat kernel trace, Pattern Recognition 42 (11) (2009) 2589–2606.

[34] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, S. Belongie, Stacked generative adversarial networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Vol. 2, 2017, p. 4.

[35] J. Yang, A. Kannan, D. Batra, D. Parikh, Lr-gan: Layered recursive generative adversarial networks for image generation, arXiv preprint arXiv:1703.01560.

[36] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4681–4690.

[37] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, in: ICML Implicit Models Workshop, 2017.

[38] Z. Zhou, J. Liang, Y. Song, L. Yu, H. Wang, W. Zhang, Y. Yu, Z. Zhang, Lipschitz generative adversarial nets, arXiv preprint arXiv:1902.05687.

[39] C. Seward, T. Unterthiner, U. Bergmann, N. Jetchev, S. Hochreiter, First order generative adversarial networks, arXiv preprint arXiv:1802.04591.

[40] J. Adler, S. Lunz, Banach wasserstein gan, in: Advances in Neural Information Processing Systems, 2018, pp. 6754–6763.

*Author Biography

Zhihong Zhang received his BSc degree (1st class Hons.) in computer science from the University of Ulster, UK, in 2009 and the PhD degree in computer science from the University of York, UK, in 2013. He won the K. M. Stott prize for best thesis from the University of York in 2013. He is now an associate professor at the software school of Xiamen University, China. His research interests are wide-reaching but mainly involve the areas of pattern recognition and machine learning, particularly problems involving graphs and networks.

Yangbin Zeng is now a master student at the school of information science and engineering of Xiamen University, China. His research interests include data mining, machine learning, and network representation learning.

Lu Bai received the Ph.D. degree from the University of York, York, UK, and both the B.Sc. and M.Sc degrees from Faculty of Information Technology, Macau University of Science and Technology, Macau SAR, P.R. China. He is now a Associate Professor in School of Information, Central University of Finance and Economics, Beijing, China. His current research interests include structural pattern recognition, machine learning, quantum walks on networks and graph matching, especially in kernel methods and complexity analysis on (hyper)graphs and networks.

Yiqun Hu received his M.D. degrees in Concord Hospital, Beijing, China, in 2006. He is now an associate professor of digestive medicine, Zhongshan Hospital, Xiamen University. His research interests include pathogenesis of Pancreatic Cancer, Pancreatic Diseases, Inflammatory Bowel Disease and Clinical application of ERCP, EUS

Meihong Wu received her PhD degree in computer science from the University of Xiamen, China, in 2009. At the same time, she is a co-educated doctor of the University of California, Los Angeles and Xiamen University. From 2009 to 2011, she was engaged in postdoctoral research in Peking University. From 2015 to 2016, she visited the University of California, Irvine as a visiting scholar. Her main research interests are cognitive computing, brain-computer fusion, audiovisual perception mechanism, auditory physiology and psychology, etc.

Shuai Wang received the Ph.D. degree from School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing,

China, in 2012, and received the B.E. degree in Measurement and Control Technology and Instrumentation from Jilin University, Changchun, China, in 2007. She is working as an Assistant Professor at Innovation Institute of Frontier Science and Technology, School of Computer Science and Engineering, Beihang University, Beijing, China. Her research interests Include Information Processing, Advanced Sensor, Artificial Intelligence, Smart City, and Information Security.

Edwin R. Hancock holds a BSc degree in physics (1977), a PhD degree in high-energy physics (1981) and a D.Sc. degree (2008) from the University of Durham, and a doctorate Honoris Causa from the University of Alicante in 2015. He is Professor in the Department of Computer Science, where he leads a group of some 25 faculty, research staff, and PhD students working in the areas of computer vision and pattern recognition. His main research interests are in the use of optimization and probabilistic methods for high and intermediate level vision. He is a fellow of the International Association for Pattern Recognition and the IEEE. He is currently Editor-in-Chief of the journal Pattern Recognition, and was founding Editor-in-Chief of IET Computer Vision from 2006 until 2012. He has also been a member of the editorial boards of the journals IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition, Computer Vision and Image Understanding, Image and Vision Computing, and the International Journal of Complex Networks.