On Computation and Generalization of Generative Adversarial Networks under Spectrum Control

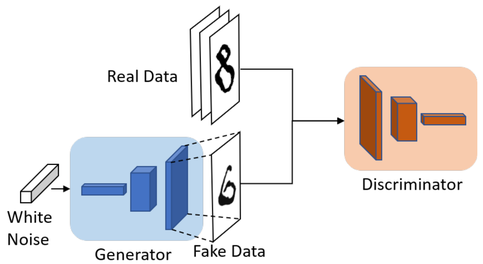
Haoming Jiang*, Zhehui Chen*, Minshuo Chen*, Feng Liu◇, Dingding Wang◇, Tuo Zhao*
*Georgia Tech ◇Florida Atlantic University

Background

GANs solve:

$$\min_{\theta}\max_{\mathcal{W}} f(\theta,\mathcal{W}) := \frac{1}{n}\sum_{i=1}^{n}\phi\left(\mathcal{A}(D_{\mathcal{W}}(x_i))\right) + \mathbb{E}_{x\sim\mathcal{D}_{G_\theta}}[\phi\left(1-\mathcal{A}(D_{\mathcal{W}}(x))\right)],$$

- G_θ denotes the **generator**.
- $D_{\mathcal{W}}$ denotes the **discriminator**.
  An L-layer discriminator can be formulated as follows:
  $$D_{\mathcal{W}}(x) = W_L\sigma_{L-1}(W_{L-1}\cdots\sigma_1(W_1x)\cdots).$$
- For DC-GAN, $\phi(x)=\log(x), \mathcal{A}(x)=\frac{1}{1+\exp(-x)}$.



SN-GAN [1] stabilizes training by spectral normalization on W_i's: $\widetilde{W}_i = \frac{W_i}{||W_i||_2}$. $||W_i||_2$ is obtained by power iteration.
Questions:

- Why does spectral normalization help?
- Can we make any further improvement?

Generalization Bound

Theorem 1. *Under some technical assumptions and assume*

- $||W_i||_2 \le B_{W_i}$ *for* $i\in[L]$ *and* $||x_k||_2 \le B_x$ *for* $k\in[n]$.
- *Generator and discriminator are well trained, i.e.,*
  $$d_{\mathcal{F},\phi}(\widehat{\mu}_n,\nu_n) - \inf_{\nu\in\mathcal{D}_G} d_{\mathcal{F},\phi}(\widehat{\mu}_n,\nu) \le \epsilon,$$
  where $d_{\mathcal{F},\phi}(\cdot,\cdot)$ *is the* ***neural distance*** *(similar to* $\mathcal{F}$*-divergence with* $\mathcal{F}$ *being the class of discriminators).*

$\Longrightarrow$ *with probability at least* $1-\delta$*, we have*

$$d_{\mathcal{F},\phi}(\mu,\nu_n) - \inf_{\nu\in\mathcal{D}_G} d_{\mathcal{F},\phi}(\mu,\nu) \le \widetilde{O}\left(\frac{B_x\prod_{i=1}^{L}B_{W_i}\sqrt{d^2L}}{\sqrt{n}}\right).$$

Spectral normalization $\Longrightarrow$ polynomial bound $\widetilde{O}\left(\sqrt{\frac{d^2L}{n}}\right)$.

Spectral normalization controls the product of spectral norms to benefit GANs in generalization.

SVD Reparameterization

$$\min_{\theta}\max_{\mathcal{W}} \quad f(\theta,\mathcal{W})$$

$\Downarrow$ SVD reparameterization: $W_i = U_iE_iV_i^\top$, where $E_i=\mathrm{diag}(e_1^i,\cdots,e_{r_i}^i)$ with $e_1^i\ge\cdots\ge e_{r_i}^i\ge 0$. Denote $\mathcal{E}=\{E_i\}_{i=1}^L, \mathcal{U}=\{U_i\}_{i=1}^L, \mathcal{V}=\{V_i\}_{i=1}^L$.

$$\min_{\theta}\max_{\mathcal{E},\mathcal{U},\mathcal{V}} \quad f(\theta,\mathcal{E},\mathcal{U},\mathcal{V}),$$
$$\text{subject to} \quad U_i^\top U_i = I_i,\quad V_i^\top V_i = I_i.$$

$\Downarrow$ Control spectrum via regularizer $\mathcal{R}(\mathcal{E})$, and constraint $\mathcal{E}\in\Omega$.

$$\min_{\theta}\max_{\mathcal{E},\mathcal{U},\mathcal{V}} \quad f(\theta,\mathcal{E},\mathcal{U},\mathcal{V}) - \gamma\mathcal{R}(\mathcal{E}),$$
$$\text{subject to} \quad \mathcal{E}\in\Omega,\quad U_i^\top U_i = I_i,\quad V_i^\top V_i = I_i.$$

$\Downarrow$ Relax orthogonal constraint for computational efficiency.

$$\min_{\theta}\max_{\mathcal{E},\mathcal{U},\mathcal{V}} f(\theta,\mathcal{E},\mathcal{U},\mathcal{V}) - \lambda\mathcal{L}_o(\mathcal{U},\mathcal{V}) - \gamma\mathcal{R}(\mathcal{E}),\ \text{s.t.}\ \mathcal{E}\in\Omega,$$
$$\text{where } \mathcal{L}_o(\mathcal{U},\mathcal{V}) = \sum_{i=1}^{L}(||U_i^\top U_i - I_i||_{\mathrm{F}}^2 + ||V_i^\top V_i - I_i||_{\mathrm{F}}^2).$$

Spectrum Control

Layer-wise Control:

$$\Omega = \left\{\mathcal{E} : 0\le e_1^i\le 1\ \forall i\in[L]\right\}.$$

- Spectral Constraint: $\widetilde{e_j^i} = \min(1, e_j^i)$.
- Spectrum Normalization: $\widetilde{e_j^i} = e_j^i/\max_j e_j^i$.

Overall Control: Alternatively, we can control the overall Lipschitz constant, $\prod_{i=1}^L e_1^i$, by Lipschitz regualzier:

$$\mathcal{R}(\mathcal{E}) := \max\left(\log\prod_{i=1}^{L} e_1^i, 0\right) = \max\left(\sum_{i=1}^{L}\log e_1^i, 0\right).$$

D-Optimal Regularizer: Penalty focusing on small singular values:

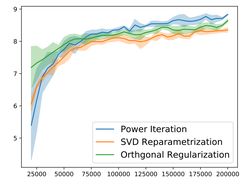
$$\mathcal{R}(\mathcal{E}) = \frac{1}{2}\sum_{i=1}^{L-1}\log\left(|(E_i^\top E_i)^{-1}|\right) = -\sum_{i=1}^{L-1}\log\left(\prod_{k=1}^{r_i} e_k^i\right).$$

Divergence Regularizer: Empirical KL divergence to a half-normal distribution p:

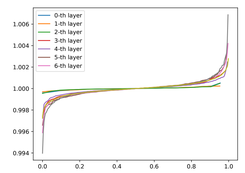
$$\mathcal{R}(\mathcal{E}) = \sum_{i=1}^{L-1}\frac{1}{r_i-1}\sum_{k=1}^{r_i-1}\log\left[\frac{(r_i-1)^{-1}}{(e_{k+1}^i - e_k^i)p(e_k^i)}\right].$$

Motivation – Slow Singular Value Decay

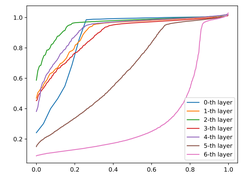
Comparison between spectral normalization (Power iteration, SVD reparametrization) and orthogonal regularizer:



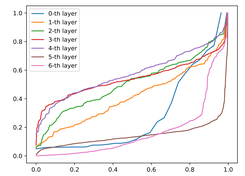
Performance



Orthogonal Regularizer
No Decay



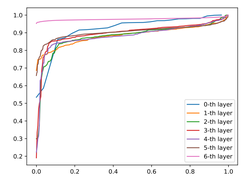
Power Iteration
Slow Decay



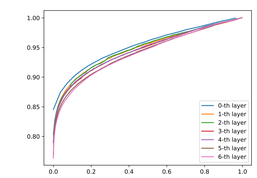
SVD Reparametrization
Fast Decay

Observation: Slow singular value decay is better than both no decay and fast decay.

Both D-Optimal and divergence regularizers yield slow singular value decay.



SN+D-Optimal Regularizer



Divergence Regularizer

Experiments

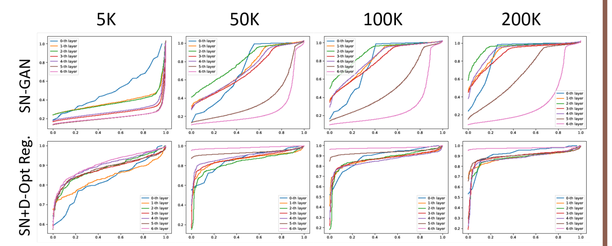
	Layer 0	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
$\|\|U^\top U-I\|\|_{\mathrm{F}}^2$	2.3e-5	1.2e-5	1.5e-5	1.6e-5	2.7e-5	2.5e-5	2.1e-5
$\|\|V^\top V-I\|\|_{\mathrm{F}}^2$	7.9e-5	1.0e-5	1.7e-5	2.5e-5	4.1e-5	7.1e-5	3.9e-5

Sub-orthogonality

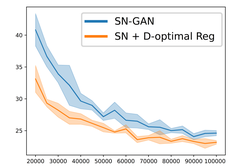
Method	Inception Score		FID	
	CIFAR-10	STL-10	CIFAR-10	STL-10
Real Data	$11.24\pm.12$	$26.08\pm.26$	7.8	7.9
CNN Baseline				
WGAN-GP	$6.72\pm.11$	$8.42\pm.09$	$39.0\pm.29$	$54.1\pm.35$
Orthogonal Reg.	$7.31\pm.09$	$8.77\pm.07$	$25.7\pm.33$	$44.5\pm.30$
SN-GAN	$7.39\pm.05$	$8.83\pm.07$	$24.7\pm.25$	$45.5\pm.34$
Ours CNN				
Spectral Norm.	$7.35\pm.05$	$8.69\pm.08$	$25.2\pm.22$	$44.8\pm.39$
Spectral Constraint	$7.43\pm.08$	$8.97\pm.05$	$24.8\pm.30$	$44.0\pm.42$
Lipschitz Reg.	$7.43\pm.08$	$8.99\pm.06$	$24.1\pm.28$	$45.3\pm.38$
SC + Divergence Reg.	$7.44\pm.05$	$9.21\pm.09$	$24.3\pm.21$	$41.9\pm.37$
SN + D-Optimal Reg.	$\mathbf{7.48\pm.06}$	$\mathbf{9.25\pm.08}$	$\mathbf{23.0\pm.27}$	$\mathbf{40.5\pm.41}$
ResNet				
Orthogonal Reg.	$7.90\pm.05$	$8.83\pm.05$	$22.3\pm.26$	$44.9\pm.35$
SN-GAN	$\mathbf{8.21\pm.05}$	$9.15\pm.06$	$\mathbf{19.5\pm.22}$	$43.0\pm.44$
SN + D-Optimal Reg.	$8.06\pm.06$	$\mathbf{9.65\pm.06}$	$20.5\pm.18$	$\mathbf{39.9\pm.33}$

Image Generation

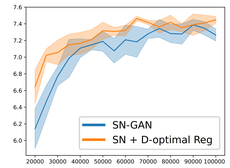
Singular Value Decay:



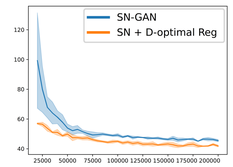
DC-GAN on CIFAR-10 and STL-10:



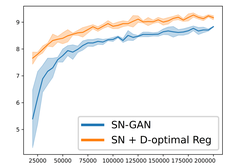
CIFAR: FID



CIFAR: Inception Score

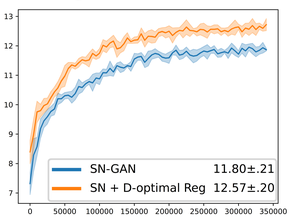


STL: FID

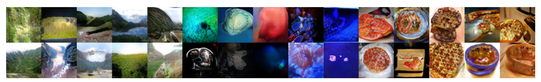


STL: Inception Score

ResNet-GAN on ImageNet:



Inception Score


Valley Jellyfish Pizza


Anemone Shoji Brain Coral


Cardoon Altar Jack-o'-lantern

Reference

[1] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.