Learning latent space representations and application to image generation

GANibal team

BENARD Maxime, BENYAMINA Mehdi, Elias Ben Rhouma

Data Science Lab 2

Dauphine PSL

November 17, 2025

Problems when training the Vanilla GAN

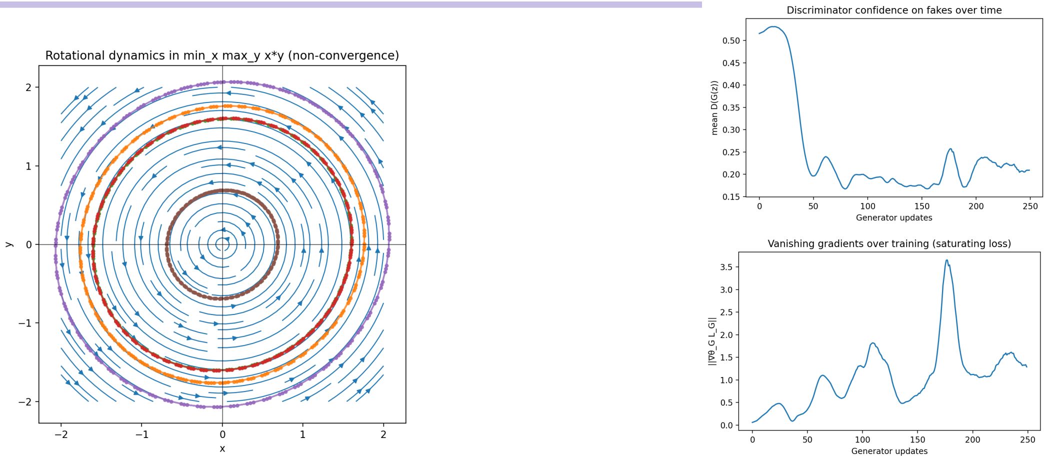

Figure: Vanilla GAN pathologies on MNIST: (left) non-convergence; (top-right) discriminator confidence; (bottom-right) generator gradient norms

Changing the divergence to Wasserstein metric

WGAN replaces the divergence with the Earth Mover distance, using the Kantorovich–Rubinstein dual

$$\mathcal{W}_1(p_{\text{data}}, p_g) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim p_{\text{data}}}[f(x)] - \mathbb{E}_{x \sim p_g}[f(x)],$$

which yields smooth, informative gradients even when supports are disjoint and induces a weaker topology (distributions converge more easily). Practically, we (i) change the losses to

$$\mathcal{L}_D = \mathbb{E}[f(G(z))] - \mathbb{E}[f(x_{\text{real}})] \quad \text{and} \quad \mathcal{L}_G = -\mathbb{E}[f(G(z))],$$

(ii) enforce 1-Lipschitzness of the critic via weight clipping, a gradient penalty or spectral normalization, and (iii) use a few more critic steps per generator update.

Approach 1: Weight Clipping (Hard Bounds)

Clipping saturation

We pick the smallest c that yields :

1. a non-trivial Wasserstein estimate $\hat{W} = \mathbb{E}[D(x_{\text{real}})] - \mathbb{E}[D(G(z))]$ (not ≈ 0),
2. stable losses (no exploding spikes),
3. low clipping saturation (20%–30% of weights at $\pm c$ over many steps).

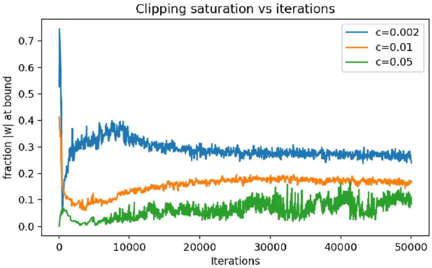

Figure: Clipping saturation versus training iterations.

Approach 2: Gradient Penalty (Soft Constraint)

WGAN-GP replaces clipping with a *gradient penalty* that directly encourages the critic to have gradient norm 1 with respect to its input:

$$L_D = \mathbb{E}_{\tilde{x}\sim p_g}\left[f_\psi(\tilde{x})\right] - \mathbb{E}_{x\sim p_{\text{data}}}\left[f_\psi(x)\right] + \lambda\,\mathbb{E}_{\hat{x}}\left(\|\nabla_{\hat{x}} f_\psi(\hat{x})\|_2 - 1\right)^2,$$
$$L_G = -\mathbb{E}_z\left[f_\psi(G_\theta(z))\right],$$

where $\hat{x} = \epsilon x + (1-\epsilon)\tilde{x}$, $x \sim p_{\text{data}}$, $\tilde{x} \sim p_g$, and $\epsilon \sim \mathcal{U}[0,1]$.

Approach 3: Spectral Normalization

Spectral normalization (SN) is another way to approximately enforce the 1-Lipschitz constraint on the critic. Instead of constraining the *outputs* of f_ψ through a penalty term, SN directly rescales each weight matrix so that its largest singular value (its *spectral norm*) is equal to 1.

For a weight matrix W, spectral normalization replaces it by

$$\bar{W} = \frac{W}{\sigma\left(W\right)}, \tag{1}$$

where $\sigma\left(W\right)$ is the spectral norm of W.

Gaussian Mixtures: A Better Latent Prior

- Standard WGAN uses $z \sim \mathcal{N}(0, I)$: **unimodal** and poorly matched to multi-modal data (digits, classes, poses).
- Replace it with a Gaussian Mixture:

$$z \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(z \mid \mu_k, \Sigma_k)$$

- **Intuition:**
  - multiple “entry points” in latent space
  - each mode can specialize (digit type, shape, style)
  - reduces mode collapse by construction
  - critic receives more diverse samples $\longrightarrow$ smoother training
- GMM aligns the latent geometry with the natural multi-modality of p_{data}.

cWGAN: Adding Conditional Structure

- Make generator and critic conditional:

$$G(z, y), \qquad f_\psi(x, y)$$

  where y = label, mixture index, or attribute.

- **Why it helps:**
  - critic compares real/fake **within each class**
  - generator no longer needs to discover classes by itself
  - reduces “global” Wasserstein difficulty into simpler subproblems
  - faster convergence, more coherent samples

- The WGAN-GP loss becomes:

$$\mathbb{E}[f(x, y)] - \mathbb{E}[f(G(z, y), y)] + \lambda(\|\nabla f\| - 1)^2$$

Putting it Together: GMM + cWGAN-GP

- Use a Gaussian Mixture prior *and* conditionality:

$$z \sim \sum_k \pi_k \mathcal{N}(z | \mu_k, \Sigma_k), \qquad G(z, y), \;\; f(x, y)$$

- **Two layers of structure:**
  - **Implicit structure (GMM):** helps generator explore multiple modes
    $\longrightarrow$ reduces collapse, improves diversity
  - **Explicit structure (conditioning):** organizes samples inside each mode
    $\longrightarrow$ sharper, more coherent outputs

- Result: more stable critic, better mode coverage, higher-quality images.

Last step: Discriminator Rejection Sampling

Goal: Improve W-GAN sample quality using discriminator-based rejection sampling.

Key Idea

- Generator = proposal distribution $p_g(x)$
- Discriminator approximates density ratio:

$$\frac{p_d(x)}{p_g(x)} \approx e^{\tilde{D}(x)}$$

Our Procedure

- Estimate maximum logit $\tilde{D}_M$
- Compute:

$$\hat{F}(x) = \tilde{D}(x) - \tilde{D}_M - \log\left(1 - e^{\tilde{D}(x) - \tilde{D}_M - \varepsilon}\right)$$

- Since tuning was unreliable, we fixed the acceptance rate to $\approx 20\%$

Conclusion

- From Vanilla GANs to WGAN-GP: progressively improved stability and gradient quality.
- WGAN-GP: better mode coverage and more reliable training than WGAN with weight clipping.
- Spectral Normalization: faster training, but weaker performance than gradient penalty.
- Gaussian Mixture priors: conceptually promising but offered no measurable improvement in practice.
- Conditional WGANs: struggled to generate coherent samples and introduced convergence difficulties when combined with GMMs.
- Discriminator Rejection Sampling (DRS): effective post-processing to filter low-quality samples.

Performance of our models

model	time(s)	FID	accuracy	recall
VanGAN	-	-	0.52	0.23
WGAN-WC	-	-	0.5	0.27
WGAN-GP	77	45	0.53	0.29
WGAN-SN (DRS)	105	52	0.5	0.44
WGAN-GP (DRS)	240	62	0.67	0.62

Table: Results

Reference

1. Arjovsky M., Chintala S., & Bottou L. *Wasserstein GAN*. Courant Institute of Mathematical Sciences & Facebook AI Research.
2. Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., & Courville A. *Improved Training of Wasserstein GANs*. Montreal Institute for Learning Algorithms (MILA), Courant Institute of Mathematical Sciences, CIFAR Fellow.
3. Mirza M., & Osindero S. *Conditional Generative Adversarial Nets*. Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, QC H3C 3J7, 2014.
4. Ben-Yosef M., & Weinshall D. *Gaussian Mixture Generative Adversarial Networks for Diverse Datasets, and the Unsupervised Clustering of Images*. School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel, 2018.
5. Azadi S., Olsson C., Darrell T., Goodfellow I., & Odena A. *Discriminator Rejection Sampling*. UC Berkeley & Google Brain.