A Variational Inequality Perspective on GANs

Gauthier Gidel*[1], **Hugo Berard***[12], Gaëtan Vignoud[1], Pascal Vincent[12], Simon Lacoste-Julien[1]

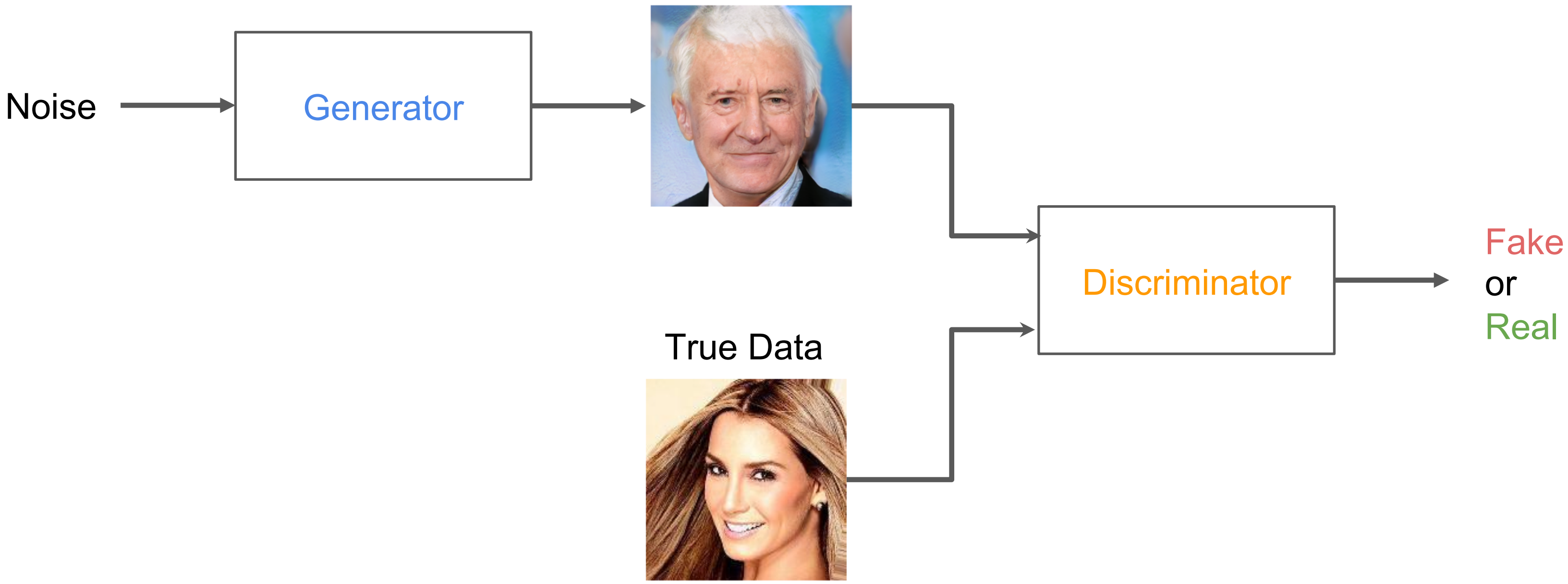
**equal contribution*

[1] MILA, Université de Montréal

[2] Facebook AI Research (FAIR), Montréal

Generative Adversarial Networks (GANs)

[Goodfelow et al. NIPS 2014]



Mila

Hugo Berard,
Montreal AI Symposium, August 28, 2018

facebook
Artificial Intelligence Research

Generative Adversarial Networks (GANs)

[Goodfelow et al. NIPS 2014]

$$\min_{\theta} \underbrace{\max_{\phi} \mathbb{E}_{x\sim p_{\mathcal{D}}}[\log(\overset{\text{Discriminator}}{\overset{\downarrow}{D_\phi}}(x))] + \mathbb{E}_{z\sim p_{\mathcal{Z}}}[\log(1 - D_\phi(\overset{\text{Generator}}{\overset{\downarrow}{G_\theta}}(z)))]}$$

If D is non-parametric: $L(\theta) = \mathrm{JSD}(p_{\mathcal{D}}||p_\theta)$

Non-saturating GAN:

$$\overbrace{\min_{\theta} -\mathbb{E}_{z\sim p_{\mathcal{Z}}}[\log(D_\phi(G_\theta(z)))]}^{\text{Loss of Generator}} \qquad \overbrace{\max_{\phi} \mathbb{E}_{x\sim p_{\mathcal{D}}}[\log(D_\phi(x))] + \mathbb{E}_{z\sim p_{\mathcal{Z}}}[\log(1 - D_\phi(G_\theta(z)))]}^{\text{Loss of Discriminator}}$$

Two-player Games

Player 1 Player 2

$$\min_{\theta} L_{\theta}(\theta, \phi) \quad \text{and} \quad \min_{\phi} L_{\phi}(\theta, \phi)$$

Zero-sum game if: $L_{\theta}(\theta, \phi) = -L_{\phi}(\theta, \phi)$ also called *Saddle Point (SP)*.

Example: WGAN formulation [Arjovsky et al. 2017]

$$\min_{\theta} \max_{\phi, ||f_{\phi}||_L \leq 1} \underbrace{\mathbb{E}_{x \sim p_{\mathcal{D}}}[f_{\phi}(x)] - \mathbb{E}_{z \sim p_{\mathcal{Z}}}[f_{\phi}(g_{\theta}(z))]}_{L_{\theta}(\theta, \phi) = -L_{\phi}(\theta, \phi)}$$

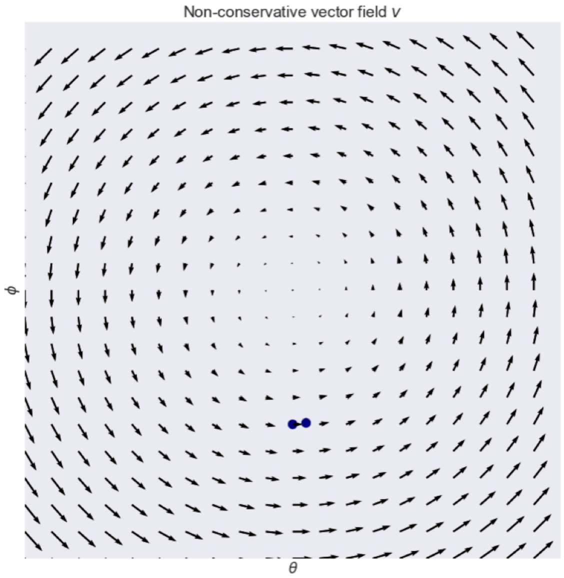
“Saddle Points are Hard to Optimize ...”

Gradient vector field: $F(\theta, \phi) = \begin{pmatrix} \nabla_\theta L_\theta(\theta, \phi) \\ \nabla_\phi L_\phi(\theta, \phi) \end{pmatrix}$

Bilinear saddle point = Linear in θ and ϕ
⇒ “Cycling behavior” (see right).

Example: WGAN with **linear** discriminator and generator

$$\min_{\theta} \max_{\phi, ||f_\phi||_L \leq 1} \phi^T \mathbb{E}_{x \sim p_\mathcal{D}}[x] - \phi^T \theta \mathbb{E}_{z \sim p_\mathcal{Z}}[z]$$



(https://www.inference.vc/my-notes-on-the-numerics-of-gans/)

… but saddle points can be optimized !

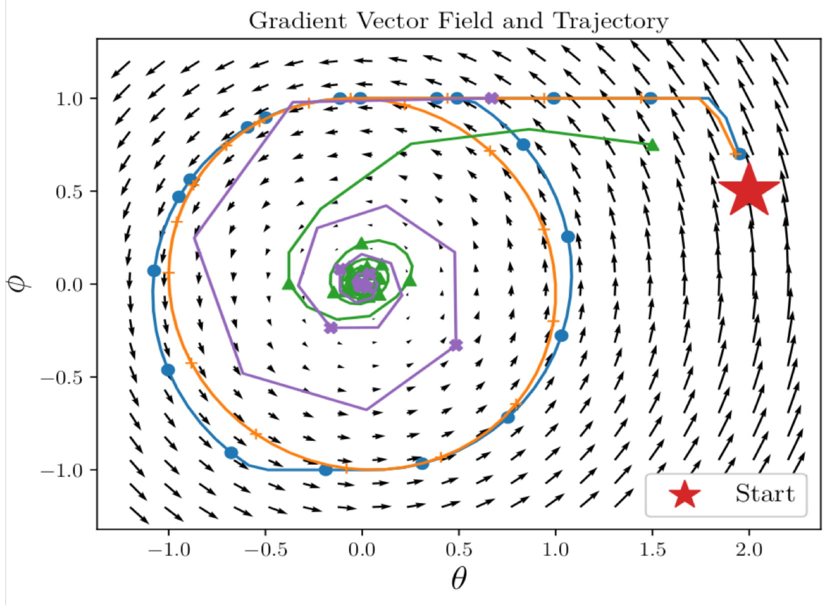
Non-convergent

- Blue: Simultaneaous gradient method.
- Orange: Alternating gradient method.

Convergent

- Green: Gradient method with **averaging**.
- Purple: **Extragradient** method.

from ***Variational Inequality*** literature



GANs as a Variational Inequality

New perspective for GANs:

- Based on **stationary conditions.**
- Relates to vast literature with standard algorithms.

Nash-Equilibrium: $\begin{cases} \theta^* = \arg\min_{\theta} L_\theta(\theta, \phi^*) \\ \phi^* = \arg\min_{\phi} L_\phi(\theta^*, \phi) \end{cases}$ ← No player can improve its cost

Stationary Conditions: $\begin{cases} \nabla_\theta L_\theta(\theta^*, \phi^*)^T (\theta - \theta^*) \geq 0 \\ \nabla_\phi L_\phi(\theta^*, \phi^*)^T (\phi - \phi^*) \geq 0 \end{cases}$ $\forall (\theta, \phi) \in \Theta \times \Phi$

can be **constraint sets**.

GANs as a Variational Inequality

Stationary Conditions: $$\begin{cases} \nabla_\theta L_\theta(\theta^*, \phi^*)^T(\theta - \theta^*) \geq 0 \\ \nabla_\phi L_\phi(\theta^*, \phi^*)^T(\phi - \phi^*) \geq 0 \end{cases} \quad \forall(\theta, \phi) \in \Theta \times \Phi$$

Can be written as: $$F(\omega) = \begin{pmatrix} \nabla_\theta L_\theta(\omega) \\ \nabla_\phi L_\phi(\omega) \end{pmatrix}$$

$\omega = (\theta, \phi)$

$$F(\omega^*)^T(\omega - \omega^*) \geq 0 \quad \forall \omega \in \Omega$$

ω^* solves the **Variational Inequality**

GANs as a Variational Inequality

Takeaways:

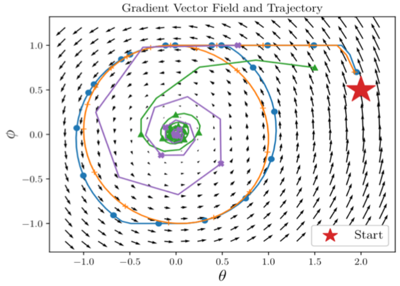
- GAN can be formulated as a **Variational Inequality.**
- Encompass <u>most</u> of GANs formulations.
- **Standard algorithms** from Variational Inequality can be used for GANs.
- **Theoretical Guarantees** (for convex and <u>*stochastic*</u> cost functions).

$$\begin{cases} \theta^* = \arg\min_{\theta} L_\theta(\theta, \phi^*) \\ \phi^* = \arg\min_{\phi} L_\phi(\theta^*, \phi) \end{cases}$$

$\downarrow$

$$F(\omega^*)^T(\omega - \omega^*) \geq 0 \quad \forall \omega \in \Omega$$

$\downarrow$



Standard Algorithms from Variational Inequality

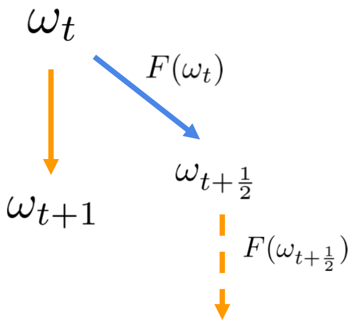
Method 1: **Averaging**

$$\bar{\omega}_T = \frac{\sum_t \rho_t \omega_t}{\sum_t \rho_t}$$

- Converge even for *"cycling behavior".*
- Easy to implement.
- Can be combined with any method.

Method 2: **Extragradient**

- Step 1: $\omega_{t+\frac{1}{2}} = \omega_t - \gamma_t F(\omega_t)$
- Step 2: $\omega_{t+1} = \omega_t - \gamma_t F(\omega_{t+\frac{1}{2}})$



Intuition: *Look 1 step in the future and anticipate next move of adversary.*

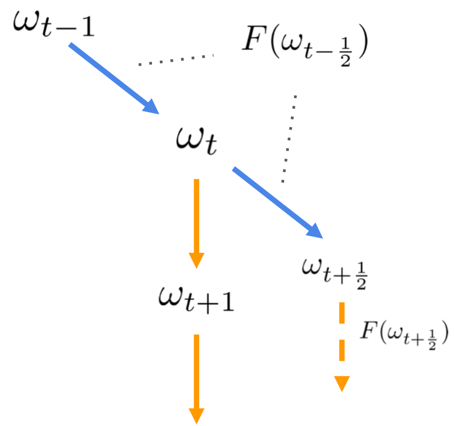
- Standard in the literature.
- Does not require *averaging.*
- *Theoretically* and *empirically* **faster.**

OneSEM: Re-using the gradients

<u>Problem</u>: Extragradient requires to compute **two** gradients at each step.

Solution: **OneSEM** ⟵ <u>Re-use</u> gradient.

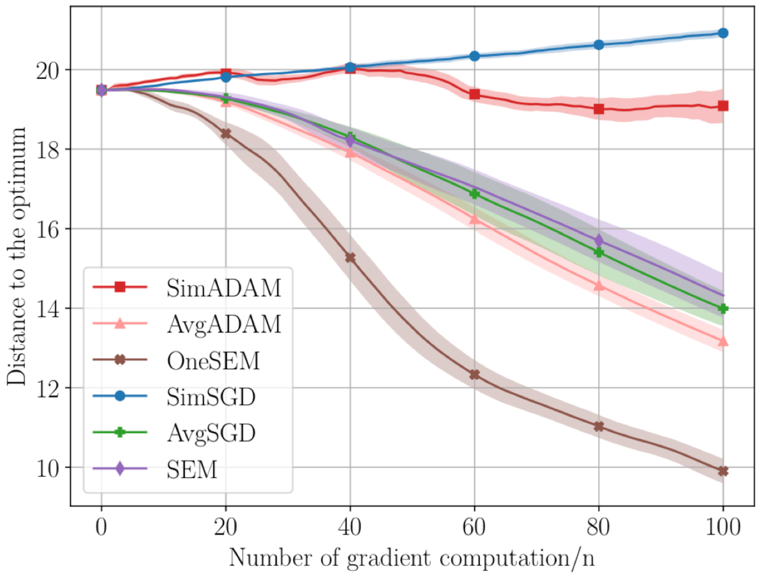
- Step 1: $\omega_{t+\frac{1}{2}} = \omega_t - \gamma_t F(\omega_{t-\frac{1}{2}})$ ⟵ ***re-use*** *from previous iteration.*
- Step 2: $\omega_{t+1} = \omega_t - \gamma_t F(\omega_{t+\frac{1}{2}})$ ⟵ (*same as* ***extragradient***).



Experimental Results

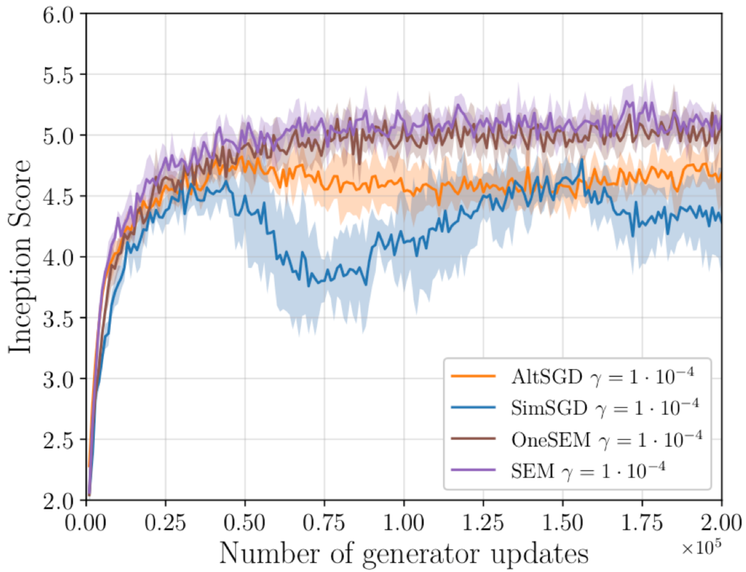
Bilinear *Stochastic* Objective:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{x}^\top \boldsymbol{M}^{(i)}\boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{a}^{(i)} + \boldsymbol{y}^\top \boldsymbol{b}^{(i)}\right).$$



Experimental Results: WGAN on CIFAR10

Inception Score vs
nb of generator updates



Inception Score on CIFAR10

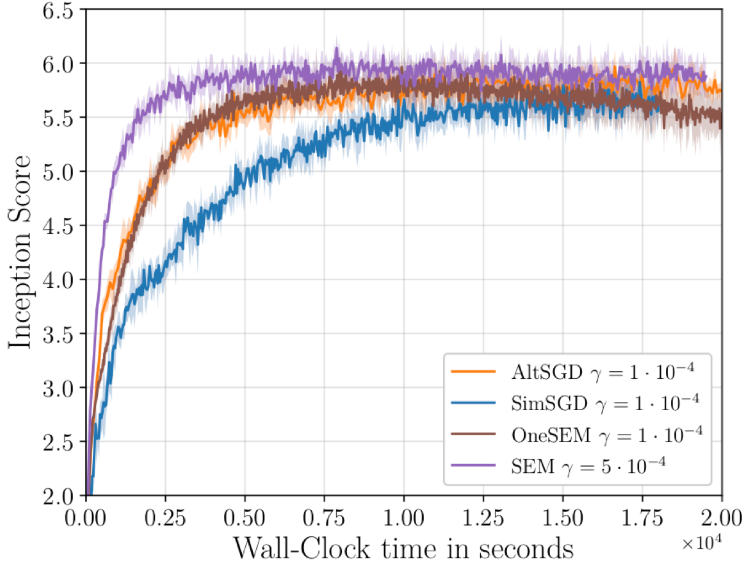
Method	no averaging	with averaging
SimSGD	5.01 ± 0.06	5.15 ± 0.15
AltSGD	5.13 ± 0.03	5.28 ± 0.08
SEM	5.52 ± 0.08	5.62 ± 0.08
OneSEM	5.45 ± 0.10	5.63 ± 0.12

(a) WGAN

Extragradient Methods

Experimental Results: WGAN-GP on CIFAR10

Inception Score vs **Time**



Inception Score on CIFAR10

Method	no averaging	with averaging
SimSGD	6.00 ± 0.07	6.01 ± 0.08
AltSGD	6.25 ± 0.05	6.51 ± 0.04
SEM	6.22 ± 0.04	6.35 ± 0.09
OneSEM	6.27 ± 0.06	6.23 ± 0.13

(b) WGAN-GP

Extragradient Methods

Conclusion

- GAN can be formulated as a **Variational Inequality.**
- Bring *standard methods* from *optimization literature* to the GAN community.
- **Averaging** helps improve the inception score (further evidence by *[Yazici et al. 2018]*).
- **Extragradient** is **faster** and achieve better convergence.
- Introduce **OneSEM** a **cheaper** version of *extragradient*.
- We can design better algorithm for GANs inspired from Variational Inequality.

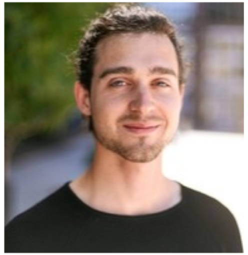
Thank you !

facebook
Artificial Intelligence Research

Université
de Montréal

Mila

Gauthier
Gidel



Gaëtan
Vignoud

Pascal
Vincent

Simon
Lacoste-Julien