

# Extragradient and Negative Momentum to Optimize GANs

Simon Lacoste-Julien

DIRO, Université de Montréal & Mila



DALI – George, South Africa – Jan 3<sup>rd</sup>, 2019

# GAN review

(slides courtesy of Aaron Courville)

# Generative modeling

- Generative models take training samples from some data distribution and learn a model that represents that distribution.

- Density estimation:



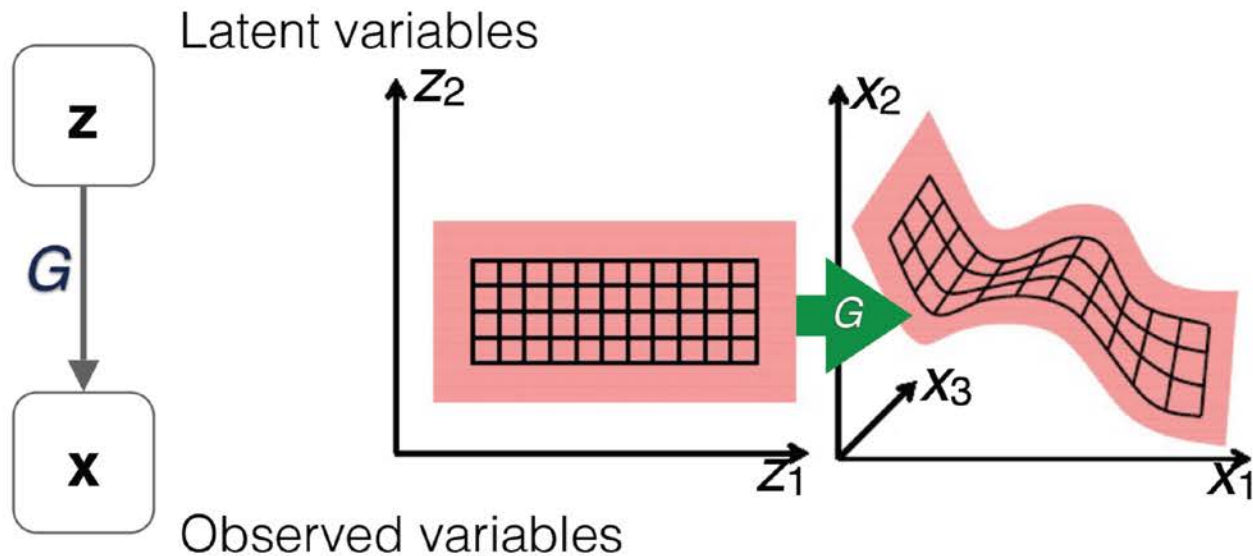
- Sample generation:



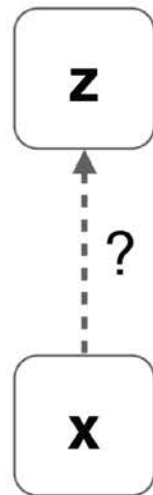
Training examples

Model samples

# How to train a latent variable generative model?

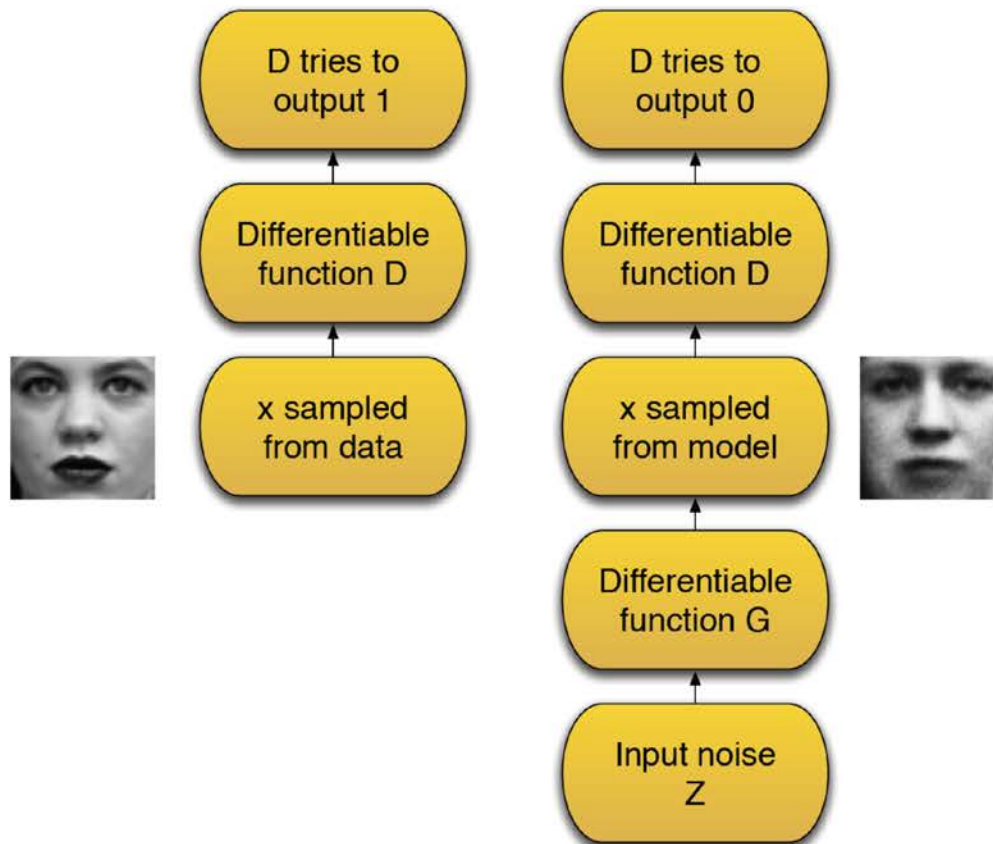


inference



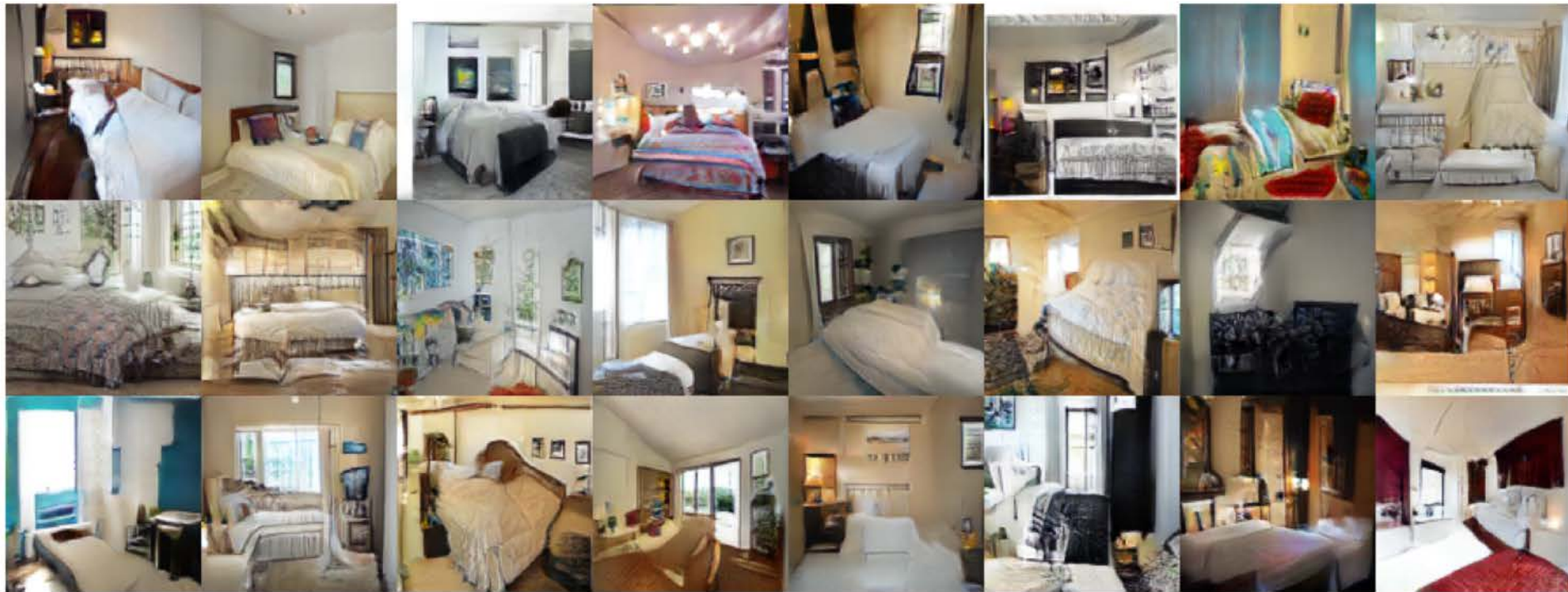
# Generative Adversarial Networks

(Goodfellow et al. 2014)



# Least-Squares GAN

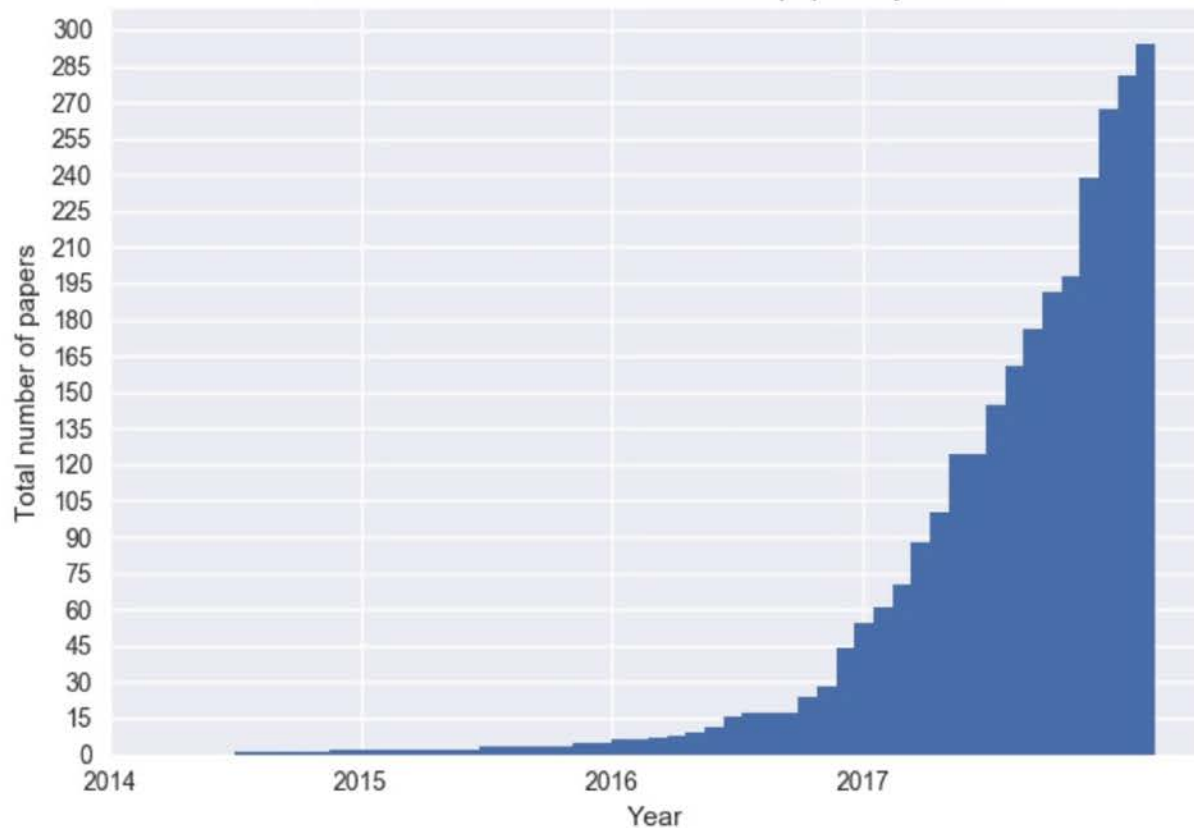
Xudong Mao, Qing Li†, Haoran Xie, Raymond Y.K. Lau and Zhen Wang, ArXiv, Feb. 2017



128x128 LSUN bedroom scenes

# An explo-GAN of papers

Cumulative number of named GAN papers by month



Explosive growth—All the named GAN variants cumulatively since 2014.

Credit: Bruno Gavranović

(from Deep Hunt, blog by Avinash Hindupur)

# Optimization for GANs

(my slides!)



# GAN and effective task-loss

---

[Goodfellow et al. NIPS 2014]

- Generative Adversarial Networks (GAN):

- generator gives  $q_\theta \in \mathcal{Q}$
- discriminator tries to distinguish  $q_\theta$  from  $p$ 
  - let  $D_\phi(x)$  be belief  $\in [0, 1]$  that  $x$  came from true distribution  $p$
- GAN (population) objective:

$$\min_{q_\theta} \max_{D_\phi} \underbrace{\mathbb{E}_{x \sim p} [\log D_\phi(x)] + \mathbb{E}_{x \sim q_\theta} [\log(1 - D_\phi(x))]}_{L_p(\theta)}$$

- if D non-parametric: get  $L_p(\theta) \approx \text{JSD}(p||q_\theta)$ 
  - (KL-like) => no good!
- but D in practice has limited capacity (e.g. NN):  $\phi \in \Phi$ 
  - => induces a meaningful task loss  $L_p(\theta)$  !
  - “parametric adversarial divergence” [Huang et al. arxiv 2017]

# A Variational Inequality Perspective on GANs

[G. Gidel\*, H. Berard\*, Gaëtan Vignoud, P. Vincent, S. Lacoste-Julien,  
arXiv 2018 – to appear at ICLR 2019]

with:



Gauthier  
Gidel



Hugo  
Berard



Pascal  
Vincent



# Optimization perspective of GAN

- Min-Max – two player (zero-sum) game:

generator  
params

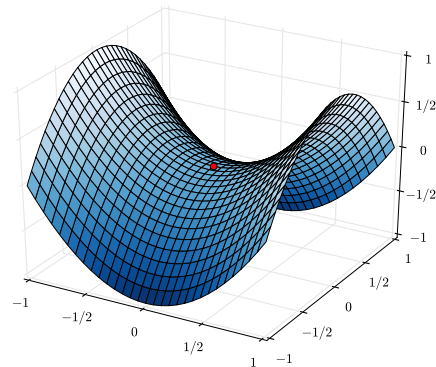
discriminator  
params

$$\min_{\theta \in \Theta} \max_{\phi \in \Phi} \bar{\mathcal{L}}(\theta, \phi).$$

- Min-max also called *saddle point* in math. programming

example: WGAN formulation:

$$\min_{\theta \in \Theta} \max_{\phi \in \Phi, \|f_{\phi}\|_L \leq 1} \mathbb{E}_{\mathbf{x} \sim p} [f_{\phi}(\mathbf{x})] - \mathbb{E}_{\mathbf{x}' \sim q_{\theta}} [f_{\phi}(\mathbf{x}')].$$



# More general...

---

as noted by [Mescheder, Nowocin, Geiger; NIPS 17]...

- Non-zero sum game: **different cost functions**

$$\begin{cases} \theta^* \in \arg \min_{\theta \in \Theta} \mathcal{L}^{(\theta)}(\theta, \phi^*) \\ \phi^* \in \arg \min_{\phi \in \Phi} \mathcal{L}^{(\phi)}(\theta^*, \phi). \end{cases}$$

$\mathcal{L}^{(\theta)} = -\mathcal{L}^{(\phi)}$  , then **zero-sum** i.e. min-max)

- Example: non-saturating GAN:

(original working version from [Goodfellow et al. NIPS 14])

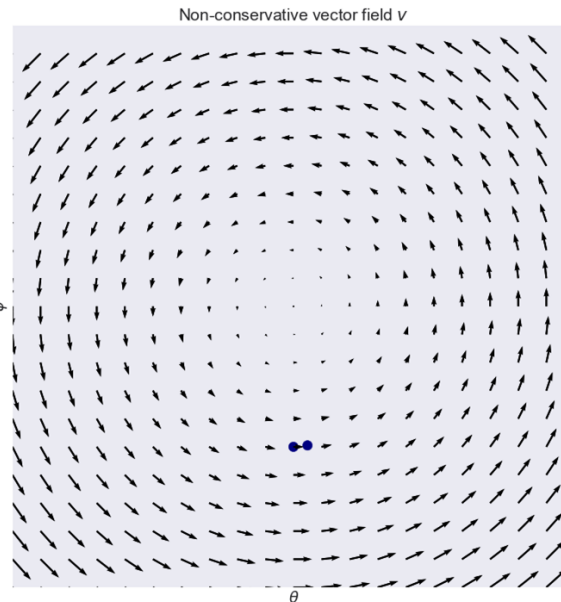
$$\mathcal{L}^{(\theta)}(\theta, \phi) \stackrel{\text{def}}{=} -\mathbb{E}_{\mathbf{x}' \sim q_{\theta}} \log f_{\phi}(\mathbf{x}')$$

$$\mathcal{L}^{(\phi)}(\theta, \phi) \stackrel{\text{def}}{=} -\mathbb{E}_{\mathbf{x} \sim p} \log f_{\phi}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}' \sim q_{\theta}} \log(1 - f_{\phi}(\mathbf{x}')).$$

# Pet peeve from GAN literature...

- “Saddle points are hard to optimize”
- example from Goodfellow NIPS 16 tutorial: *simple bilinear objective*
- WGAN with *linear* discriminator and generator:

$$\min_{\theta \in \Theta} \max_{\phi \in \Phi, \|\phi\| \leq 1} \phi^T \mathbb{E}[\mathbf{X}] - \phi^T \theta \mathbb{E}[\mathbf{Z}].$$



# Standard methods from Math. Prog.

- But saddle points can be optimized!

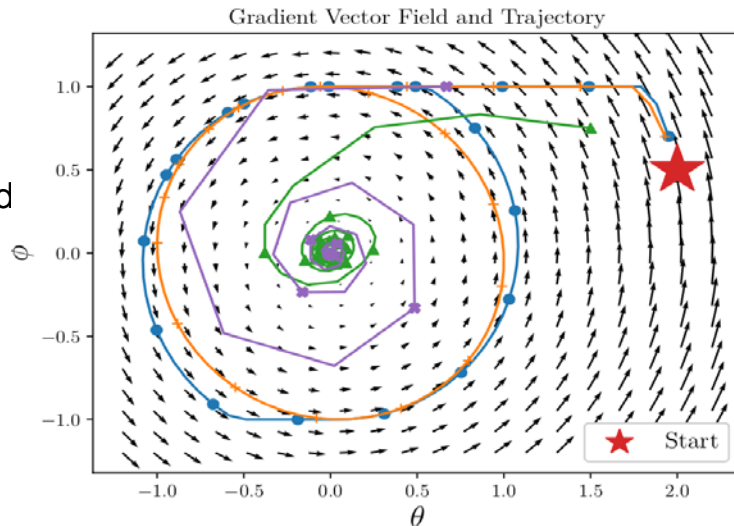
**non-convergent**

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

---

**convergent**

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



from *variational inequality* literature

# GAN as a variational inequality

---

- new perspective on GANs
  - based on stationary conditions
  - relate to vast literature in math. opt. with standard algorithms

- stationary conditions for game:

$$\begin{cases} \theta^* \in \mathbf{arg\,min}_{\theta \in \Theta} \mathcal{L}^{(\theta)}(\theta, \phi^*) \\ \phi^* \in \mathbf{arg\,min}_{\phi \in \Phi} \mathcal{L}^{(\phi)}(\theta^*, \phi). \end{cases}$$

$$\begin{cases} \nabla_{\theta} \mathcal{L}^{(\theta)}(\theta^*, \phi^*)^{\top} (\theta - \theta^*) \geq 0, \\ \nabla_{\phi} \mathcal{L}^{(\phi)}(\theta^*, \phi^*)^{\top} (\phi - \phi^*) \geq 0 \end{cases} \quad \forall (\theta, \phi) \in \Theta \times \Phi.$$

- interpretation: each player can't improve cost function

# GAN as a variational inequality

---

- stationary conditions:

$$\begin{cases} \nabla_{\theta} \mathcal{L}^{(\theta)}(\theta^*, \phi^*)^{\top}(\theta - \theta^*) \geq 0, \\ \nabla_{\phi} \mathcal{L}^{(\phi)}(\theta^*, \phi^*)^{\top}(\phi - \phi^*) \geq 0 \end{cases} \quad \forall (\theta, \phi) \in \Theta \times \Phi.$$

- rewrite as:

Defining  $\mathbf{x} \stackrel{\text{def}}{=} (\theta, \phi)$ ,  $\mathbf{x}^* \stackrel{\text{def}}{=} (\theta^*, \phi^*)$ ,  $\mathcal{X} \stackrel{\text{def}}{=} \Theta \times \Phi$  and,

$$F(\mathbf{x}) \stackrel{\text{def}}{=} \begin{pmatrix} \nabla_{\theta} \mathcal{L}^{(\theta)}(\theta, \phi) \\ \nabla_{\phi} \mathcal{L}^{(\phi)}(\theta, \phi) \end{pmatrix}, \quad (7)$$

the stationary conditions can be compactly formulated as:

$$F(\mathbf{x}^*)^{\top}(\mathbf{x} - \mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}. \quad (8)$$

say " $\mathbf{x}^*$  solves the **variational inequality**"



# GAN as a variational inequality

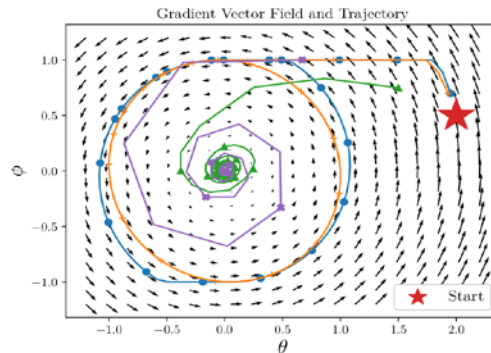
takeaways:

- GAN objective even with different cost functions for the discriminator and the generator,
- Can be formulated as a **variational inequality** .
- Can import to GANs standard variational inequality algorithms (+ make their stochastic extensions)
- **With theoretical guarantees (at least for “convex” cost functions)**

$$\begin{cases} \theta^* \in \arg \min_{\theta \in \Theta} \mathcal{L}^{(\theta)}(\theta, \phi^*) \\ \phi^* \in \arg \min_{\phi \in \Phi} \mathcal{L}^{(\phi)}(\theta^*, \phi). \end{cases}$$



$$F(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}.$$



# Convergent methods for V.I.

---

## gradient method with averaging

[Bruck 1977]

stochastic: [Nemirovski et  
al. 09]

## extragradient method

[Korpelich 1976]

stochastic: [Juditsky et al.  
11]

- smaller variance

---

### Algorithm 1 Stochastic Gradient Method (AvgSGD)

---

- 1: Let  $x_0 \in \mathcal{X}$
  - 2: **for**  $t = 0 \dots T - 1$  **do**
  - 3:   Sample  $\xi_t \sim P$
  - 4:   Update  $\mathbf{x}_{t+1} := P_{\mathcal{X}}(\mathbf{x}_t - \gamma_t F(\mathbf{x}_t, \xi_t))$
  - 5: **end for**
  - 6: Return  $\bar{\mathbf{x}}_T = \sum_{t=0}^{T-1} \gamma_t \mathbf{x}_t / \sum_{t=0}^{T-1} \gamma_t$
- 

---

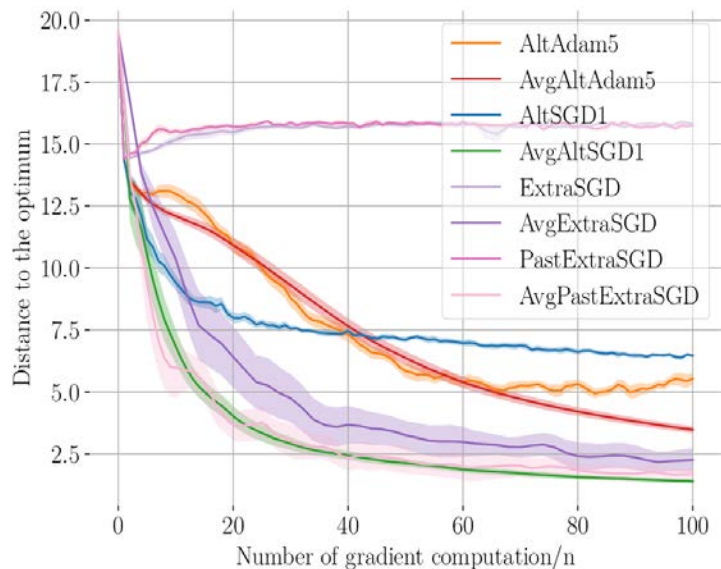
### Algorithm 2 Stochastic Extragradient Method (SEM)

---

- 1: Let  $x_0 \in \mathcal{X}$
  - 2: **for**  $t = 0 \dots T - 1$  **do**
  - 3:   Sample  $\xi_t \sim P$   $\triangleright$  *Extrapolation step*
  - 4:   Compute  $\mathbf{x}_{t+\frac{1}{2}} := P_{\mathcal{X}}(\mathbf{x}_t - \gamma_t F(\mathbf{x}_t, \xi_t))$
  - 5:   Sample  $\xi_{t+\frac{1}{2}} \sim P$   $\triangleright$  *Update step*
  - 6:   Update  $\mathbf{x}_{t+1} := P_{\mathcal{X}}(\mathbf{x}_t - \gamma_t F(\mathbf{x}_{t+\frac{1}{2}}, \xi_{t+\frac{1}{2}}))$
  - 7: **end for**
  - 8: Return  $\bar{\mathbf{y}}_T = \sum_{t=0}^{T-1} \gamma_t \mathbf{x}_{t+1/2} / \sum_{t=0}^{T-1} \gamma_t$
-

# Bilinear stochastic experiment

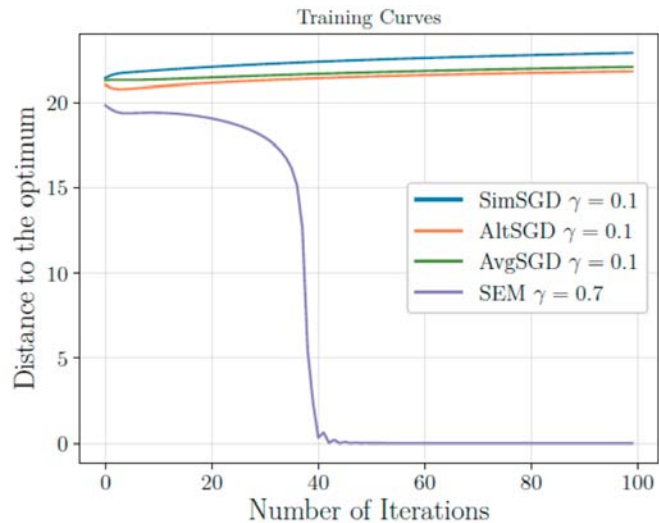
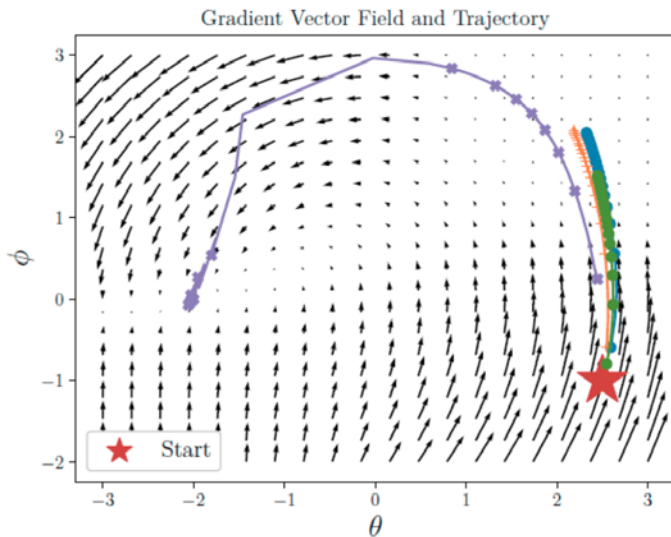
bilinear stochastic objective:  $\frac{1}{n} \sum_{i=1}^n \left( \mathbf{x}^\top \mathbf{M}^{(i)} \mathbf{y} + \mathbf{x}^\top \mathbf{a}^{(i)} + \mathbf{y}^\top \mathbf{b}^{(i)} \right)$ .



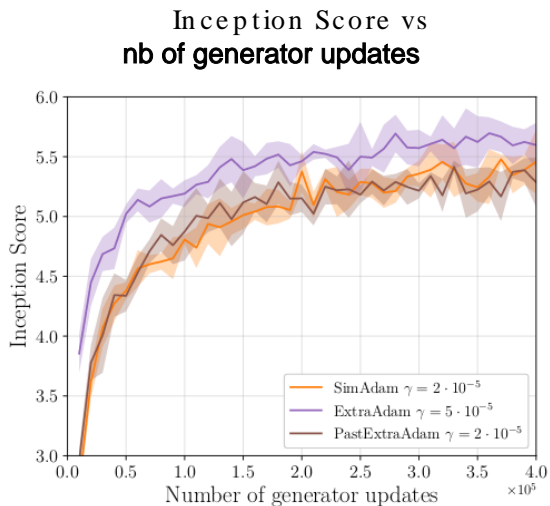
# Non-convex toy example

- min-max GAN with:
  - linear discriminator  $D_\phi(\mathbf{x}) = \phi^T \mathbf{x}$
  - $p$  &  $q$  are diracs  $p = \delta_{\mathbf{x}_0}$   $q_\theta = \delta_\theta$

$$\min_{\theta} \max_{\phi} -\log(1 + e^{-\phi^T \theta}) - \log(1 + e^{\phi^T \mathbf{x}_0})$$



# Experimental Results: WGAN on CIFAR10



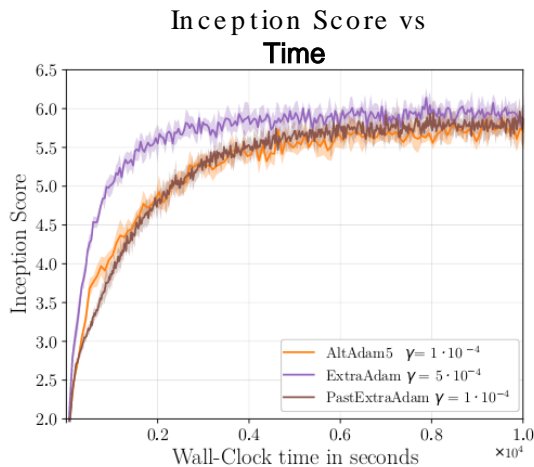
Inception Score on CIFAR10

Model	WGAN		
	no averaging	uniform avg	EMA
SimAdam	$6.05 \pm .12$	$5.83 \pm .16$	$6.08 \pm .10$
AltAdam5	$5.45 \pm .08$	$5.72 \pm .06$	$5.49 \pm .05$
ExtraAdam	<b><math>6.38 \pm .09</math></b>	<b><math>6.38 \pm .20</math></b>	<b><math>6.37 \pm .08</math></b>
PastExtraAdam	$5.98 \pm 0.15$	$6.07 \pm 0.19$	$6.01 \pm 0.11$
OptimAdam	$5.74 \pm 0.10$	$5.80 \pm 0.08$	$5.78 \pm 0.05$

↑  
Extragradient Methods

↑  
Averaging

# Experimental Results: WGAN -GP on CIFAR 10



Inception Score on CIFAR 10

Model	WGAN-GP		
	no averaging	no averaging	uniform avg
SimAdam	$6.05 \pm .12$	$6.00 \pm .07$	$6.01 \pm .08$
AltAdam5	$5.45 \pm .08$	$6.25 \pm .05$	<b><math>6.51 \pm .05</math></b>
ExtraAdam	<b><math>6.38 \pm .09</math></b>	$6.22 \pm .04$	$6.35 \pm .05$
PastExtraAdam	$5.98 \pm 0.15$	$6.27 \pm 0.06$	$6.23 \pm 0.13$
OptimAdam	$5.74 \pm 0.10$	-	-

↑  
Extragradient Methods

↑  
Averaging

# Latest results

---

Model	WGAN (DCGAN)			WGAN-GP (ResNet)	
Method	no averaging	uniform avg	EMA	no averaging	uniform avg
SimAdam	$6.05 \pm .12$	$5.83 \pm .16$	$6.08 \pm .10$	$7.54 \pm .21$	$7.74 \pm .27$
AltAdam5	$5.45 \pm .08$	$5.72 \pm .06$	$5.49 \pm .05$	$7.20 \pm .06$	$7.67 \pm .15$
ExtraAdam	<b><math>6.38 \pm .09</math></b>	<b><math>6.38 \pm .20</math></b>	<b><math>6.37 \pm .08</math></b>	$7.79 \pm .09$	<b><math>8.26 \pm .12</math></b>
PastExtraAdam	$5.98 \pm .15$	$6.07 \pm .19$	$6.01 \pm .11$	$7.71 \pm .12$	$7.84 \pm .18$
OptimAdam	$5.74 \pm .10$	$5.80 \pm .08$	$5.78 \pm .05$	$7.80 \pm .07$	$7.99 \pm .12$

# Discussion

---

- New framework called variational inequality to consider GAN objectives.
- Tapped into the optimization literature to bring new techniques to train GANs
- With caveats:
  - Averaging seems to consistently improve the inception score at convergence,
  - Extragradient looks to be more stable for step size tuning.
- Just considered two standard algorithms.
- Could design algorithm specific to GANs inspired from variational inequalities.



# Negative Momentum for Improved Game Dynamics

Gauthier Gidel, Reyhane Askari Hemmat, Mohammad Pezeshki, Gabriel Huang, Remi Lepriol, Simon Lacoste-Julien, Ioannis Mitliagkas

<https://arxiv.org/abs/1807.04740>

to appear at AISTATS 2019

Game dynamics are ~~weird~~

fascinating

Start with optimization

d y n a m i c s

# Optimization

$$\theta^* \in \arg \min_{\theta \in \Theta} \mathcal{L}^{(\theta)}(\theta)$$

Smooth, differentiable cost function,  $L$

→ Looking for stationary (fixed) points  
(gradient is 0)

→ Gradient descent

# Optimization

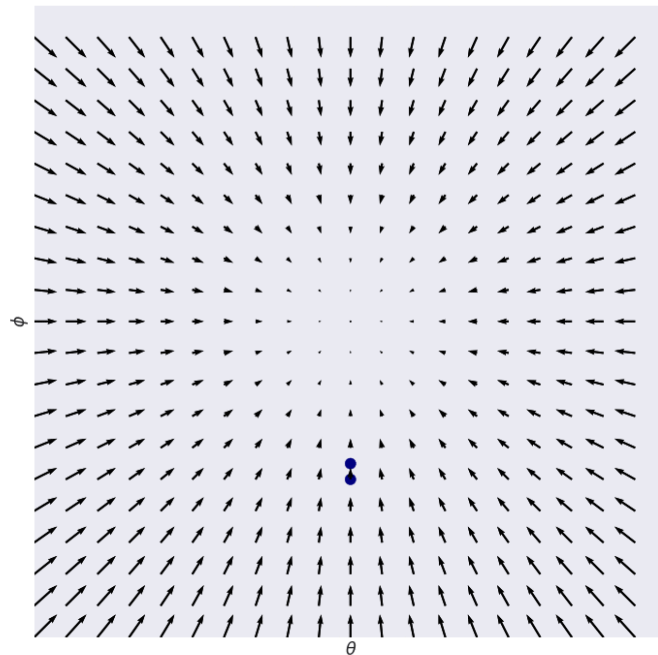
$$\mathbf{v}(\boldsymbol{\theta}) = \nabla \mathcal{L}^{(\boldsymbol{\theta})}(\boldsymbol{\theta})$$

Conservative vector field



Straightforward dynamics

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \mathbf{v}(\boldsymbol{\theta}_t)$$



# Gradient descent

$$\mathbf{v}(\boldsymbol{\theta}) = \nabla \mathcal{L}^{(\boldsymbol{\theta})}(\boldsymbol{\theta})$$

Conservative vector field

→

Straightforward dynamics

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \mathbf{v}(\boldsymbol{\theta}_t)$$

Fixed-point analysis

$$F_\eta(\boldsymbol{\theta}) = \boldsymbol{\theta} - \eta \mathbf{v}(\boldsymbol{\theta})$$

Jacobian of operator

$$\nabla F_\eta(\boldsymbol{\theta}) = I - \eta \nabla \mathbf{v}(\boldsymbol{\theta})$$

**Hessian of objective, L**

# Local convergence

**Theorem 1** (Prop. 4.4.1 Bertsekas [1999]). *If the spectral radius  $\rho_{\max} \stackrel{\text{def}}{=} \rho(\nabla F_{\eta}(\omega^*)) < 1$ , then, for  $\omega_0$  in a neighborhood of  $\omega^*$ , the distance of  $\omega_t$  to the stationary point  $\omega^*$  converges at a linear rate of  $\mathcal{O}((\rho_{\max} + \epsilon)^t)$ ,  $\forall \epsilon > 0$ .*

Eigenvalues of op. Jacobian

$$\lambda(\nabla F_{\eta}(\boldsymbol{\theta})) = 1 - \eta \lambda(\nabla \mathbf{v}(\boldsymbol{\theta}))$$

If  $\rho(\boldsymbol{\theta}^*) = \max |\lambda(\boldsymbol{\theta}^*)| < 1$ , then  
fast local convergence

Jacobian of operator

$$\nabla F_{\eta}(\boldsymbol{\theta}) = I - \eta \nabla \mathbf{v}(\boldsymbol{\theta})$$

**Hessian of objective, L**  
**Symmetric, real-eigenvalues**

# Momentum (heavy ball, Polyak 1964)

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \mathbf{v}(\boldsymbol{\theta}_t) + \beta(\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})$$

Jacobian of momentum operator

$$\nabla F_{\eta, \beta}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{I}_n & \mathbf{0}_n \end{bmatrix} - \eta \begin{bmatrix} \nabla \mathbf{v}(\boldsymbol{\theta}_t) & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} + \beta \begin{bmatrix} \mathbf{I}_n & -\mathbf{I}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}$$

**Non-symmetric, with complex eigenvalues**  
**→ Rotations in augmented state-space**



Games

# Games

Nash Equilibrium

$$\theta^* \in \arg \min_{\theta \in \Theta} \mathcal{L}^{(\theta)}(\theta, \varphi^*)$$

$$\varphi^* \in \arg \min_{\varphi \in \Phi} \mathcal{L}^{(\varphi)}(\theta^*, \varphi)$$

Smooth, differentiable  $L$   
→ Looking for local Nash eq

→ Gradient descent

→ **Simultaneous**

→ **Alternating**

# Game dynamics

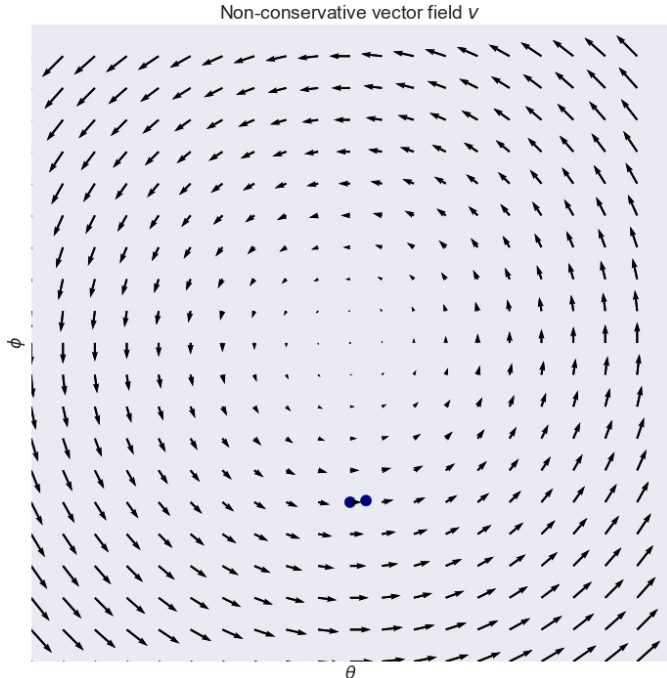
$$v(\varphi, \theta) := \begin{bmatrix} \nabla_{\varphi} \mathcal{L}^{(\varphi)}(\varphi, \theta) \\ \nabla_{\theta} \mathcal{L}^{(\theta)}(\varphi, \theta) \end{bmatrix}$$

Non-conservative vector field

→

Rotational dynamics

$$F_{\eta}(\varphi, \theta) \stackrel{\text{def}}{=} [\varphi \quad \theta]^{\top} - \eta v(\varphi, \theta)$$



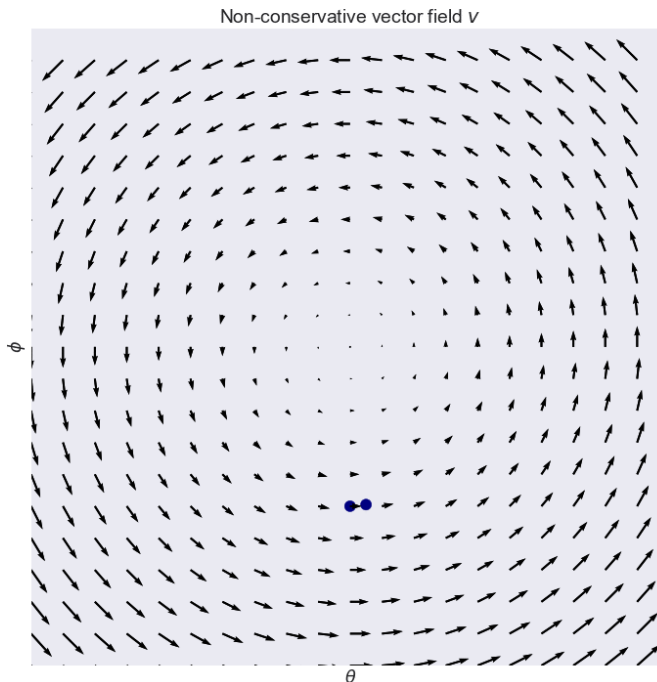
# Game dynamics under gradient descent

$$F_{\eta}(\varphi, \theta) \stackrel{\text{def}}{=} [\varphi \quad \theta]^{\top} - \eta v(\varphi, \theta)$$

**Jacobian is non-symmetric, with complex eigenvalues  $\rightarrow$  Rotations in decision space**

Games already demonstrate rotational dynamics.

Can momentum help/hurt??



# Spoiler

Positive momentum can be bad for adversarial games

Practice that was very common when GANs were first invented:

- Recent work reduced the momentum.
- Not an accident

# Our results

Negative momentum is “optimal” on simple bilinear game

Negative momentum values are locally preferable near 0 on a more general class of games

Negative momentum is empirically best for certain zero sum games like “saturating GANs”

# Momentum on games

Recall Polyak's momentum (on top of simultaneous grad. desc.):

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \eta \mathbf{v}(\mathbf{x}_t) + \beta(\mathbf{x}_t - \mathbf{x}_{t-1}), \quad \mathbf{x}_t = (\boldsymbol{\theta}_t, \phi_t)$$

Fixed point operator requires a **state augmentation** :  
(because we need previous iterate)

$$F_{\eta, \beta}(\mathbf{x}_t, \mathbf{x}_{t-1}) := \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{I}_n & \mathbf{0}_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{bmatrix} - \eta \begin{bmatrix} \mathbf{v}(\mathbf{x}_t) \\ \mathbf{0}_n \end{bmatrix} + \beta \begin{bmatrix} \mathbf{I}_n & -\mathbf{I}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{bmatrix}$$

# Bilinear game

$$\min_{\theta} \max_{\varphi} \theta^{\top} A \varphi$$

Method	$\beta$	Bounded	Converges
Simultaneous	$\beta \in \mathbb{R}$	$\times$	$\times$
Alternated	$>0$	$\times$	$\times$
	$0$	$\checkmark$	$\times$
	$<0$	$\checkmark$	$\checkmark$



# “Proof by picture”

Gradient descent



Simultaneous

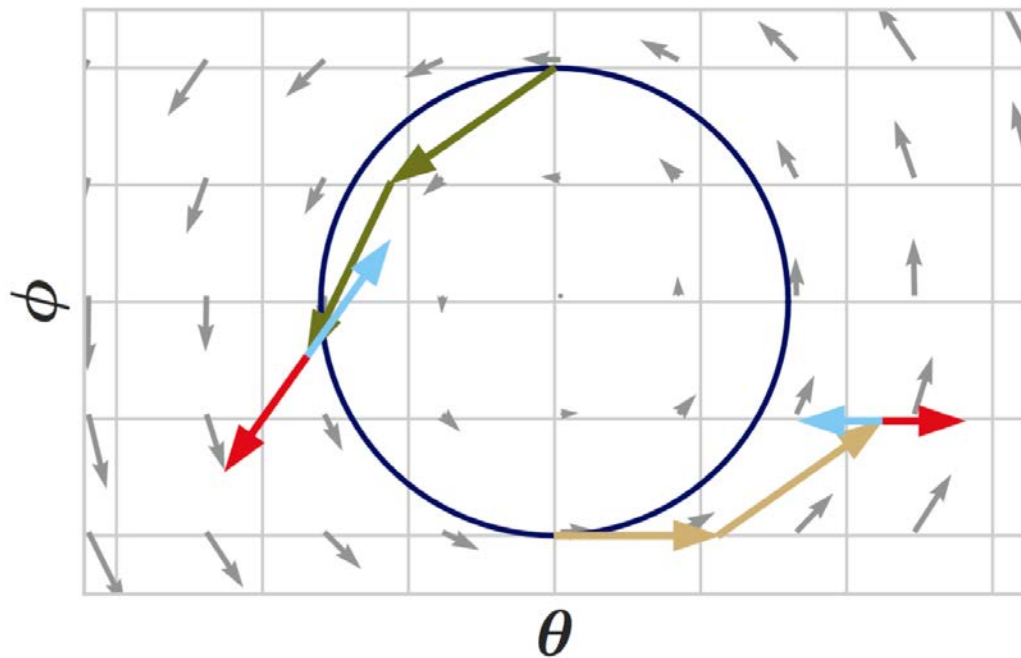


Alternating

Momentum

→ Positive

→ Negative



# General games

## Without momentum

$$F_\eta(\boldsymbol{\varphi}, \boldsymbol{\theta}) \stackrel{\text{def}}{=} [\boldsymbol{\varphi} \quad \boldsymbol{\theta}]^\top - \eta \mathbf{v}(\boldsymbol{\varphi}, \boldsymbol{\theta})$$

$$\nabla F_\eta(\boldsymbol{\theta}) = I - \eta \nabla \mathbf{v}(\boldsymbol{\theta})$$

$$\text{Sp}(\nabla F_\eta(\boldsymbol{\omega}^*)) = \{1 - \eta\lambda \mid \lambda \in \text{Sp}(\nabla \mathbf{v}(\boldsymbol{\omega}^*))\}$$

## With momentum

$$F_{\eta,\beta}(\boldsymbol{\omega}_t, \boldsymbol{\omega}_{t-1}) = (\boldsymbol{\omega}_{t+1}, \boldsymbol{\omega}_t)$$

$$\boldsymbol{\omega}_{t+1} := \boldsymbol{\omega}_t - \eta \mathbf{v}(\boldsymbol{\omega}_t) + \beta(\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1})$$

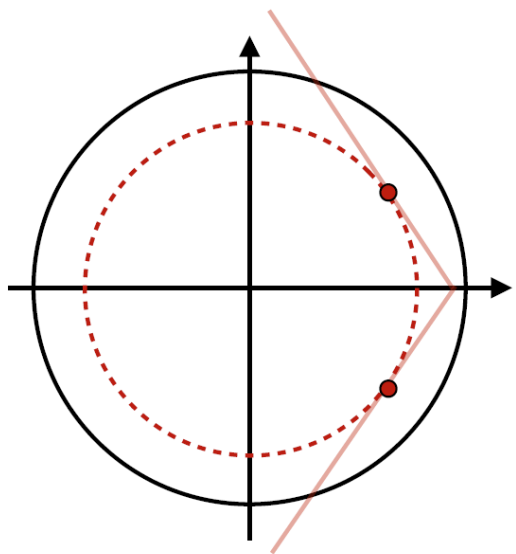
$$\begin{bmatrix} I_n & \mathbf{0}_n \\ I_n & \mathbf{0}_n \end{bmatrix} - \eta \begin{bmatrix} \nabla \mathbf{v}(\boldsymbol{\omega}_t) & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} + \beta \begin{bmatrix} I_n & -I_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}$$

$$\mu_\pm(\beta, \eta, \lambda) := (1 - \eta\lambda + \beta) \frac{1 \pm \Delta^{\frac{1}{2}}}{2},$$

Shifting the eigen-values

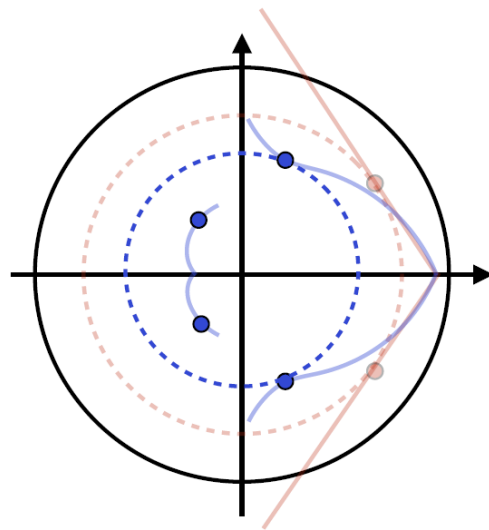
Theorem 4: Given a large class of games, a small negative momentum can improve the magnitude of the limiting eigenvalues

Without momentum



$$\text{Sp}(\nabla F_\eta(\omega^*)) = \{1 - \eta\lambda \mid \lambda \in \text{Sp}(\nabla v(\omega^*))\}$$

With negative momentum

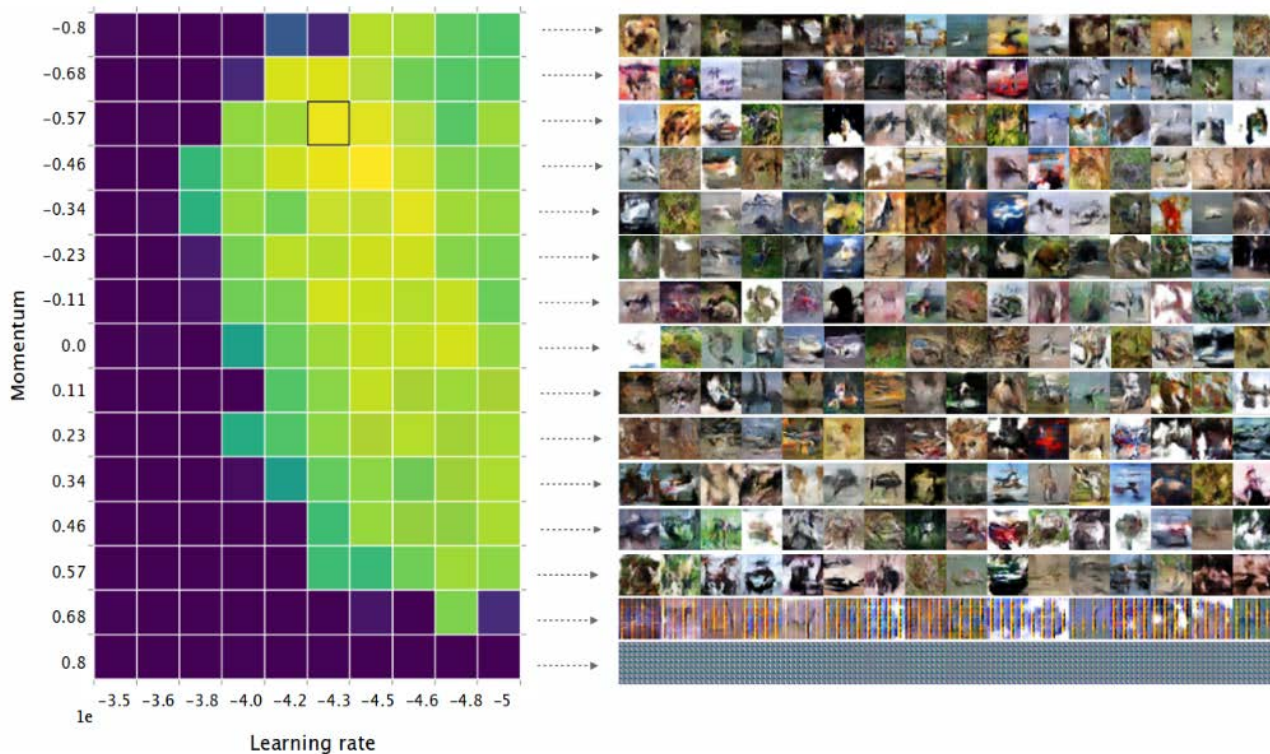


$$\mu_{\pm}(\beta, \eta, \lambda) := (1 - \eta\lambda + \beta) \frac{1 \pm \Delta^{\frac{1}{2}}}{2},$$

# Empirical results

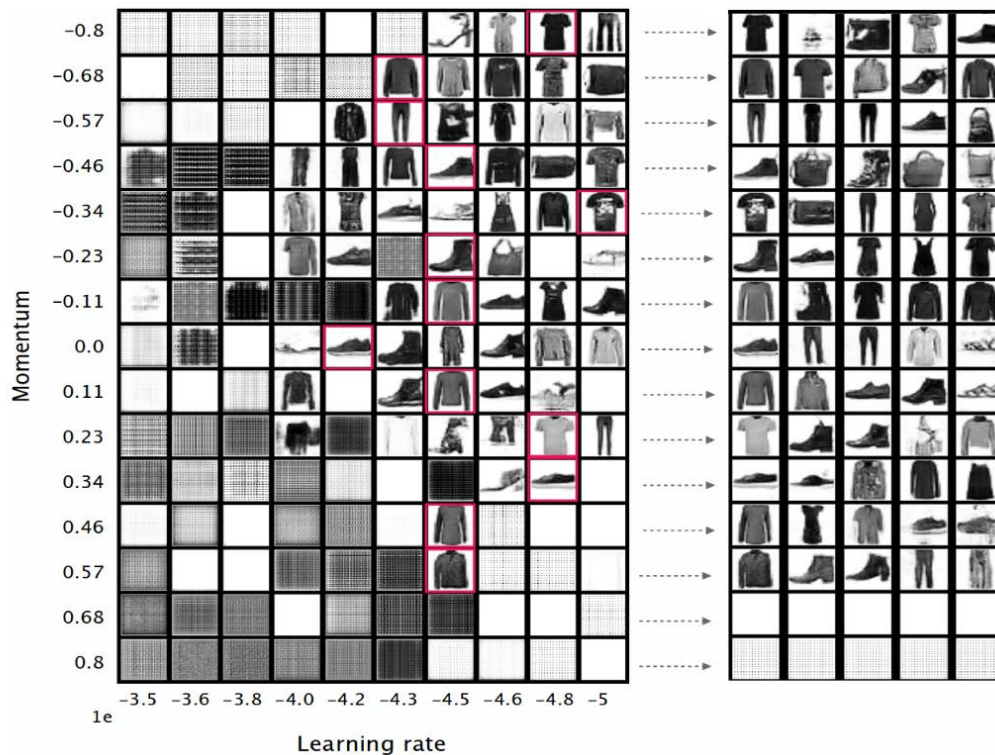
# What happens in practice ?

CIFAR-10 :



# What happens in practice ?

Fashion MNIST:



# Negative Momentum

To sum up:

- Negative momentum seems to improve the behaviour due to “bad” eigenvalues.
- Optimal for a class of games
- Empirically optimal on “saturating” GANs



# SMOOTH GAMES OPTIMIZATION AND MACHINE LEARNING WORKSHOP

Room 512 ABEF, Friday Dec 7th,  
NeurIPS2018 , Montreal.

VIDEOS OF THE WORKSHOP

<https://sgo-workshop.github.io/>

Thank you !