Dynamics of Generative Adversarial Network in the high-dimensional limit

Chuang Wang （王闯）
wangchuang@ia.ac.cn

Institute of **A**utomation, **C**hinese **A**cademy of **S**cience (CASIA)

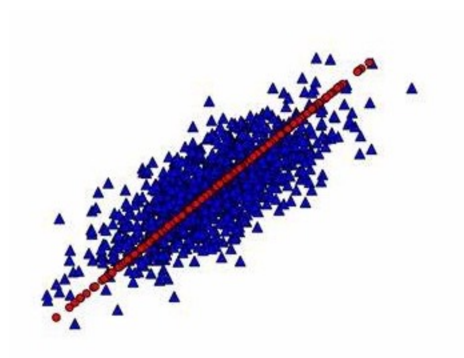
July 30, 2019 @ CSRC-ICTP Joint Workshop on Big Data, Machine Learning and Complexity Research

Unsupervised generative models

Principal Component Analysis

High Dimensions → Low Dimensions

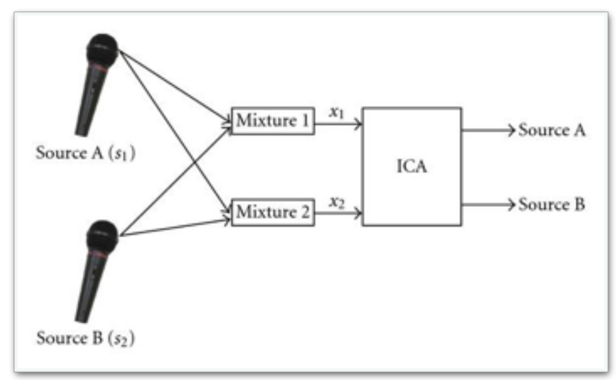
$$\boldsymbol{A}\boldsymbol{z} \to \boldsymbol{x}$$
$$\boldsymbol{z} \sim \mathcal{N}$$



Independent Component Analysis

Coupled signal → Decoupled

$$\boldsymbol{a}_1 z_1 + \boldsymbol{a}_2 z_2 \to \boldsymbol{x}$$
$$\text{Assume: } z_1 \perp z_2$$

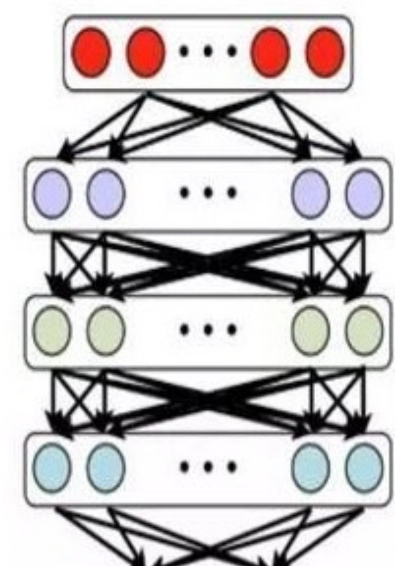


Generative Adversarial Net

Simple variables →

Complex variables

$$f_m(\ldots f_2(f_1(\boldsymbol{z}))\ldots) \to \boldsymbol{x}$$



- Objective: Retrieve semantic meaningful features
- General Idea: Use generative model to approximate the real-world data.

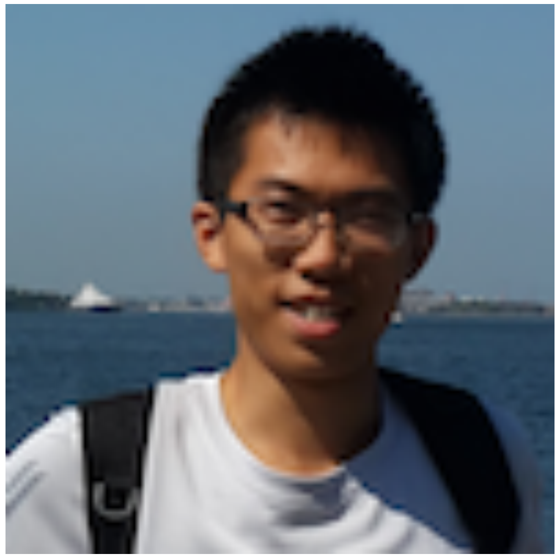
[1] **Chuang Wang**, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

[2] **Chuang Wang**, Yonina C. Eldar, Yue M. Lu, Subspace Estimation from Incomplete Observations: A High-Dimensional Analysis, IEEE Journal of Selected Topics in Signal Processing, 2018

[3] **Chuang Wang**, Yue M. Lu, The scaling limit of high-dimensional online independent component analysis, NIPS 2017

Collaborators

Yue Lu @ Harvard



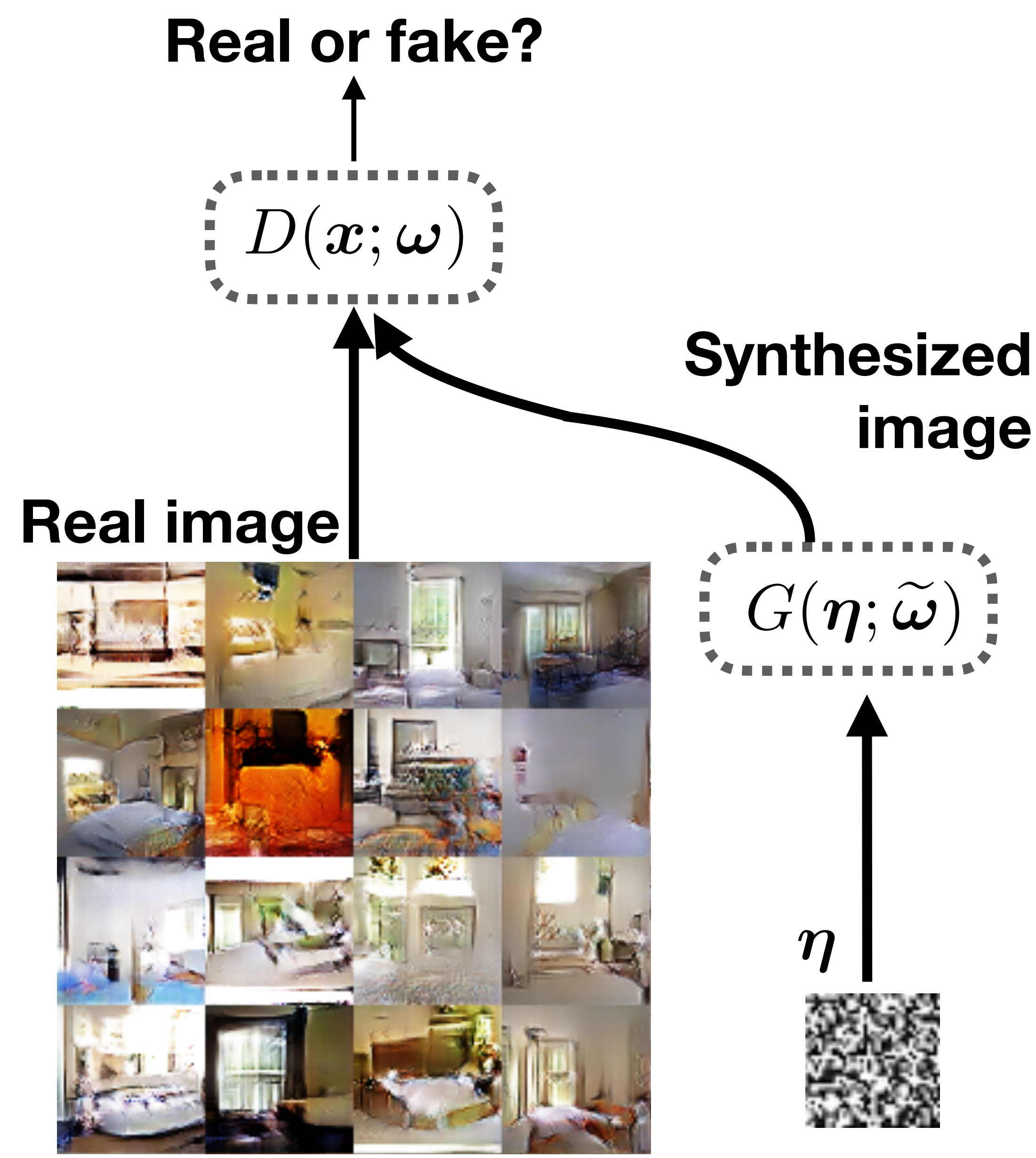
Hong Hu @ Harvard

Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Introduction of GAN

GAN: sample from an unknown distribution [Goodfellow2014]

- **G**enerator
- **D**iscriminator



Training: two-player game

- **D** aims to distinguish whether an input is real or comes from **G**
- **G** aims to fool **D**

Generative Adversarial Network (GAN)

Progressive GAN, Nvidia, ICLR 2018

Generative Adversarial Network (GAN)

- **Challenge**
  - **Multiple stationary points**
  - **Oscillation**
  - **Mode Collapsing**
- **Objective**： $\min_{\boldsymbol{w}^G} \max_{\boldsymbol{w}^D} \mathbb{E}_{\boldsymbol{y}^R} \mathbb{E}_{\boldsymbol{y}^G \sim p(\boldsymbol{w}^G)} J(\boldsymbol{y}^R, \boldsymbol{y}^G; \boldsymbol{w}^D, \boldsymbol{w}^G)$

$$J(\boldsymbol{y}^R, \boldsymbol{y}^G; \boldsymbol{w}^D, \boldsymbol{w}^G) = D_1(\boldsymbol{y}^R; \boldsymbol{w}^D) - D_2(\boldsymbol{y}^G; \boldsymbol{w}^D) + \text{regularizer}$$

Probability that the input is real

- **Training method:**

Stochastic gradient descent/ascent

$$\boldsymbol{w}_{k+1}^{\mathrm{D}} = \boldsymbol{w}_k^{\mathrm{D}} + \tau^{\mathrm{D}} \nabla_{\boldsymbol{w}_k^D} J$$
$$\boldsymbol{w}_{k+1}^{\mathrm{G}} = \boldsymbol{w}_k^{\mathrm{G}} - \tau^{\mathrm{G}} \nabla_{\boldsymbol{w}_k^G} J$$

Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Precise analysis in high-dimensions

Main idea: Stochastic learning as a stochastic dynamics

Stochastic Learning ······▶ Performance optimization
Adaptive parameters

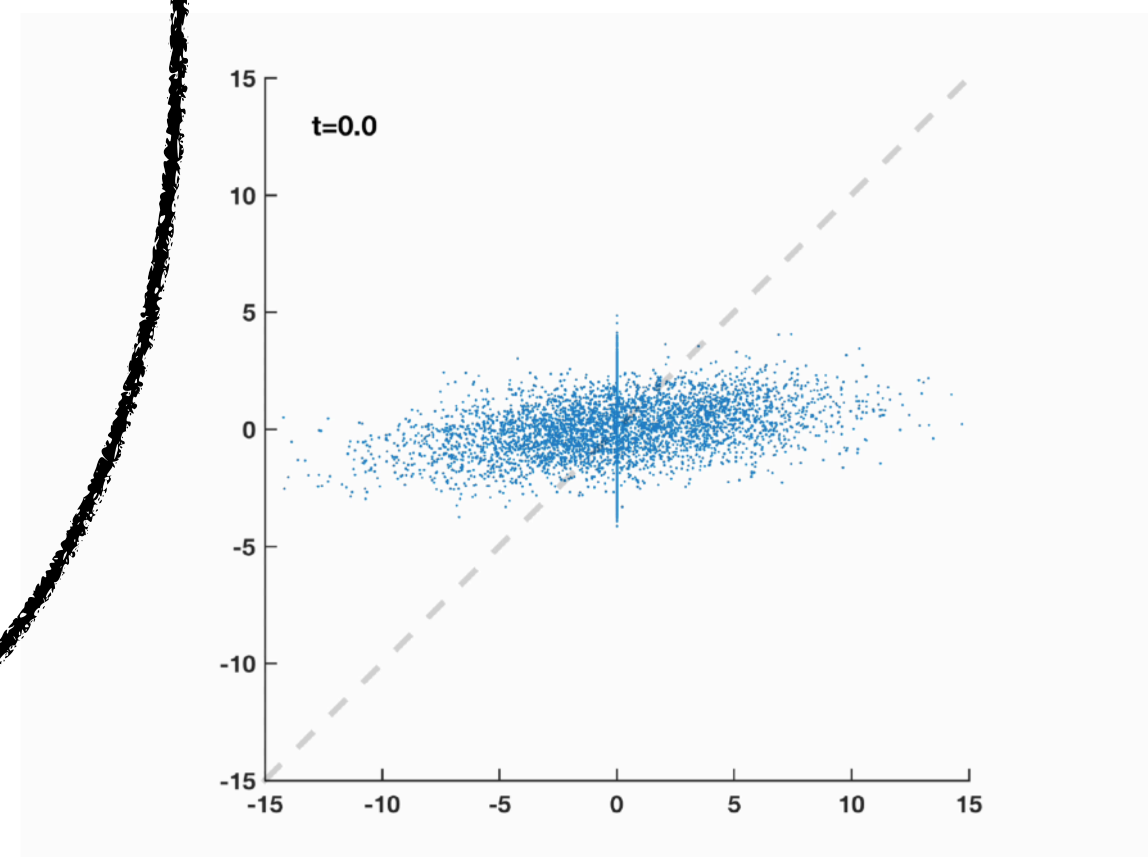
Differential equations

$$\mathrm{d}X_t = a(X_t)\mathrm{d}t + b(X_t)\mathrm{d}B_t$$

Stochastic modeling

Analytical/numerical solution
Fix points analysis

Performance Prediction



Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

A simple solvable GAN model

Real data: $$\boldsymbol{y}_k^R = g(\boldsymbol{\xi} c_k + \boldsymbol{a}_k)$$

Generator: $$\boldsymbol{y}_k^G = \widetilde{g}(\boldsymbol{w}^G \widetilde{c}_k + \widetilde{\boldsymbol{a}}_k)$$

Discriminator: $$D(\boldsymbol{y}; \boldsymbol{w}^D) = \widehat{D}(\boldsymbol{y}^T \boldsymbol{w}^D)$$

Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Training algorithm

- **Training algorithm**
  Stochastic gradient ascent/descent

$$\boldsymbol{w}_{k+1}^{\mathrm{D}} = \boldsymbol{w}_k^{\mathrm{D}} + \tau^{\mathrm{D}} \nabla_{\boldsymbol{w}_k^D} J$$
$$\boldsymbol{w}_{k+1}^{\mathrm{G}} = \boldsymbol{w}_k^{\mathrm{G}} - \tau^{\mathrm{G}} \nabla_{\boldsymbol{w}_k^G} J$$

- **Small learning rate analysis** $\frac{\tau^{\mathrm{D}}}{\tau^{\mathrm{G}}} = \alpha,\ \tau^{\mathrm{D}} \to 0$ $\quad n$ is finite

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}_t^{\mathrm{D}} = \mathbb{E}\left[\nabla_{\boldsymbol{w}_t} J\right]$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}_t^{\mathrm{G}} = -\alpha\mathbb{E}\left[\nabla_{\boldsymbol{w}_t} J\right]$$

Mescheder et al., 2017; Nagarajan & Kolter, 2017;
Roth et al., 2017; Mescheder et al., 2018;
Heusel et al., 2017; Li et al., 2017
E V. Mazumdar, M. I. Jordan, S. S. Sastry, 2019

- **High-dimensional analysis** $\frac{\tau^{\mathrm{D}}}{\tau^{\mathrm{G}}} = \alpha,\ n\tau^D = \tau,\ \tau^{\mathrm{D}} \to 0,\ i.e.\ n \to \infty$

$$\mathrm{d}\boldsymbol{w}_t^{\mathrm{D}} = \widetilde{\tau}^{\mathrm{D}}\mathbb{E}[\nabla_{\boldsymbol{w}_t^{\mathrm{D}}} J]\mathrm{d}t + \widetilde{\tau}^{\mathrm{D}}\sqrt{\mathrm{var}[\nabla_{\boldsymbol{w}_t^{\mathrm{D}}} J]}\mathrm{d}B_t$$
$$\mathrm{d}\boldsymbol{w}_t^{\mathrm{G}} = -\widetilde{\tau}^{\mathrm{G}}\mathbb{E}[\nabla_{\boldsymbol{w}_t^{\mathrm{G}}} J]\mathrm{d}t + \widetilde{\tau}^{\mathrm{G}}\sqrt{\mathrm{var}[\nabla_{\boldsymbol{w}_t^{\mathrm{G}}} J]}\mathrm{d}\widetilde{B}_t$$

 Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Dynamics of Microscopic state

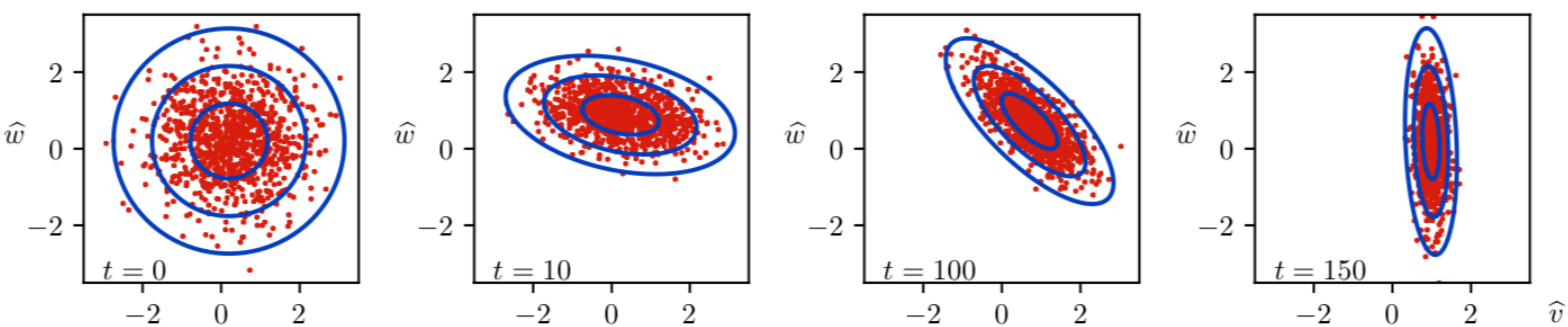
Dynamics of gradient flow in High-dimensional limit

$$\mathrm{d}w_t^{\mathrm{G}} = \widetilde{\tau}(\widetilde{g}_t w_t^{\mathrm{D}} + L_t w_t^{\mathrm{G}})\mathrm{d}t$$
$$\mathrm{d}w_t^{\mathrm{D}} = \tau(g_t\, \xi + \widetilde{g}_t w_t^{\mathrm{G}} + h_t w_t^{\mathrm{D}})\mathrm{d}t + \tau\sqrt{b_t}\mathrm{d}B_t$$

where B_t is the standard Brownian motion,

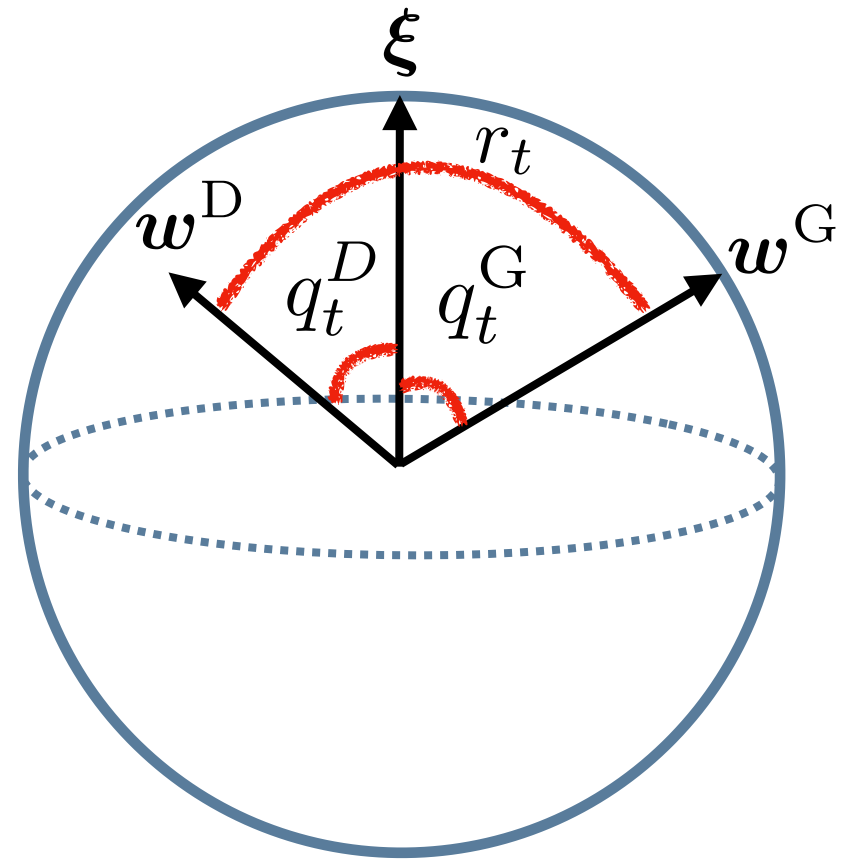
and g_t, $\widetilde{g}_t$, L_t, h_t and b_t are some deterministic functions.

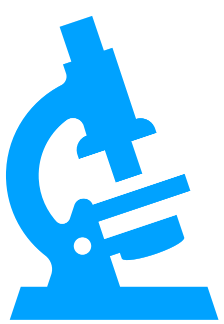
Probability law: Integral partial differential equation



 Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Microscopic states and Macroscopic state





Microscopic states

$\boldsymbol{\xi}$ $\boldsymbol{w}^{\mathrm{G}}$ $\boldsymbol{w}^{\mathrm{D}}$

Three n-D vectors

Macroscopic states

q_t^D q_t^{G} r_t

Three scalars

 Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Main Theory on Macroscopic dynamics

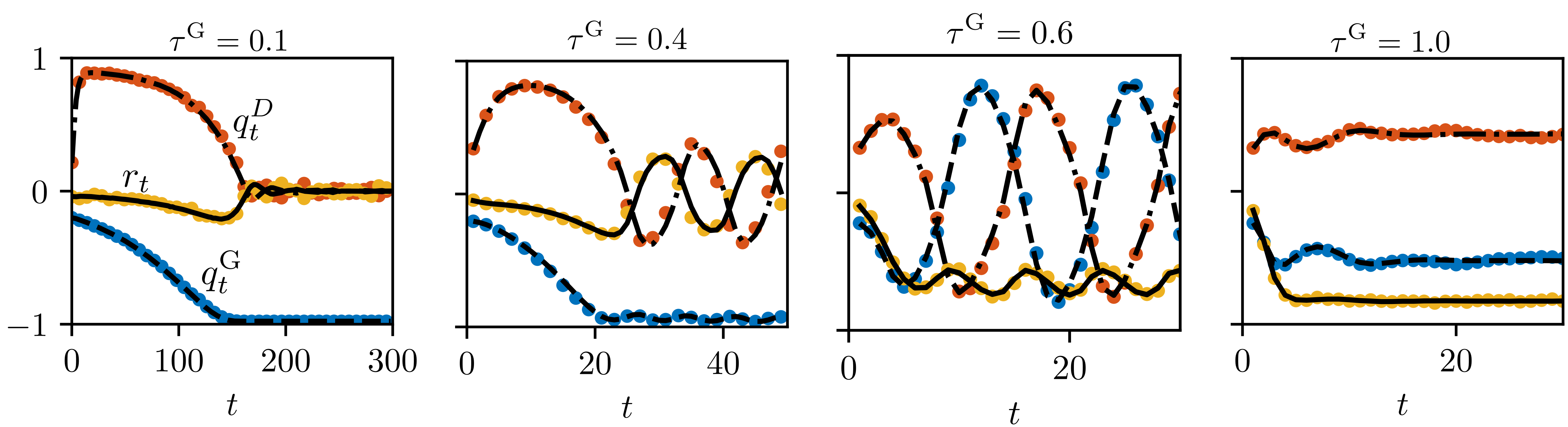
Rescaled time $t = k/n$

Theorem: As $n \to \infty$, $(q_t^{\mathrm{D}}, q_t^{\mathrm{G}}, r_t)$ converges weakly to the unique solution of the system of ODEs

$$\frac{d}{dt}(q_t^D, q_t^G, r_t) = \boldsymbol{g}(q_t^D, q_t^G, r_t)$$

Rigorous characterization:

$$\max_{0 \le k \le nT} \mathbb{E}[|q_t^{\mathrm{G}} - q^{\mathrm{G},n}(\tfrac{k}{n})| + |q_t^{\mathrm{D}} - q^{\mathrm{D},n}(\tfrac{k}{n})| + |r_t - r^n(\tfrac{k}{n})|] \le \frac{C(T)}{\sqrt{n}}$$



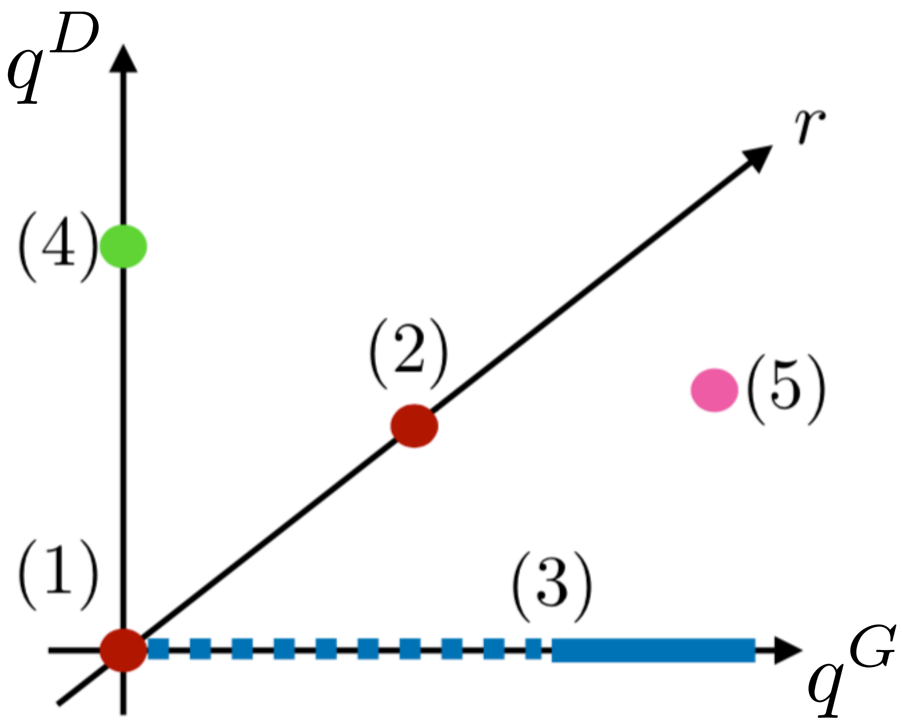
[1] C. Wang, H. Hu, Y. M. Lu, A Solvable High-Dimensional Model of GAN, NIPS, 2019

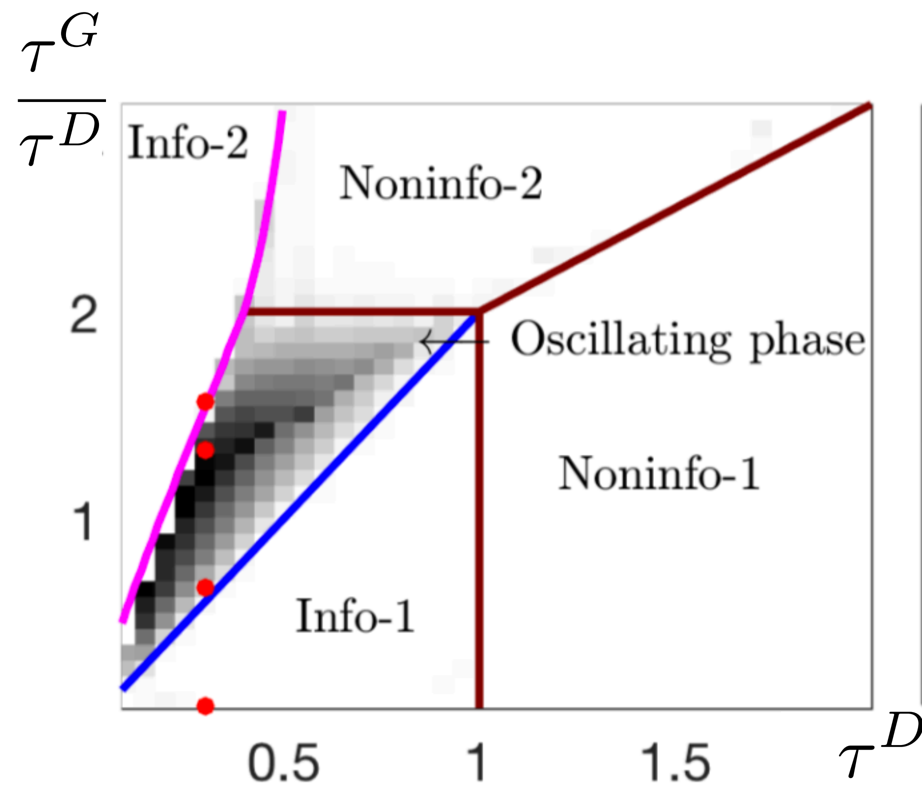
[2] C. Wang, Y. C. Eldar, Y. M. Lu, Subspace Estimation from Incomplete Observations: A High-Dimensional Analysis, IEEE JSTSP, 2018

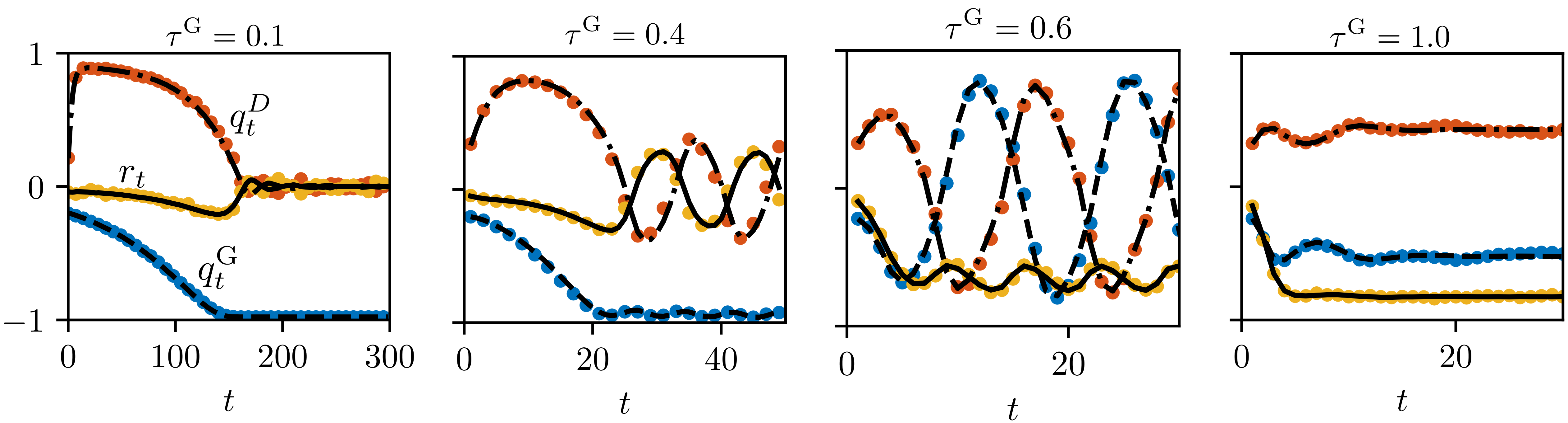
[3] S. Goldt, M. Advani, A. M. Saxe, F. Krzakala, and L. Zdeborová, "Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup," NIPS 2019

Stationary State analysis

$$\frac{d}{dt}(q_t^D, q_t^G, r_t) = 0$$



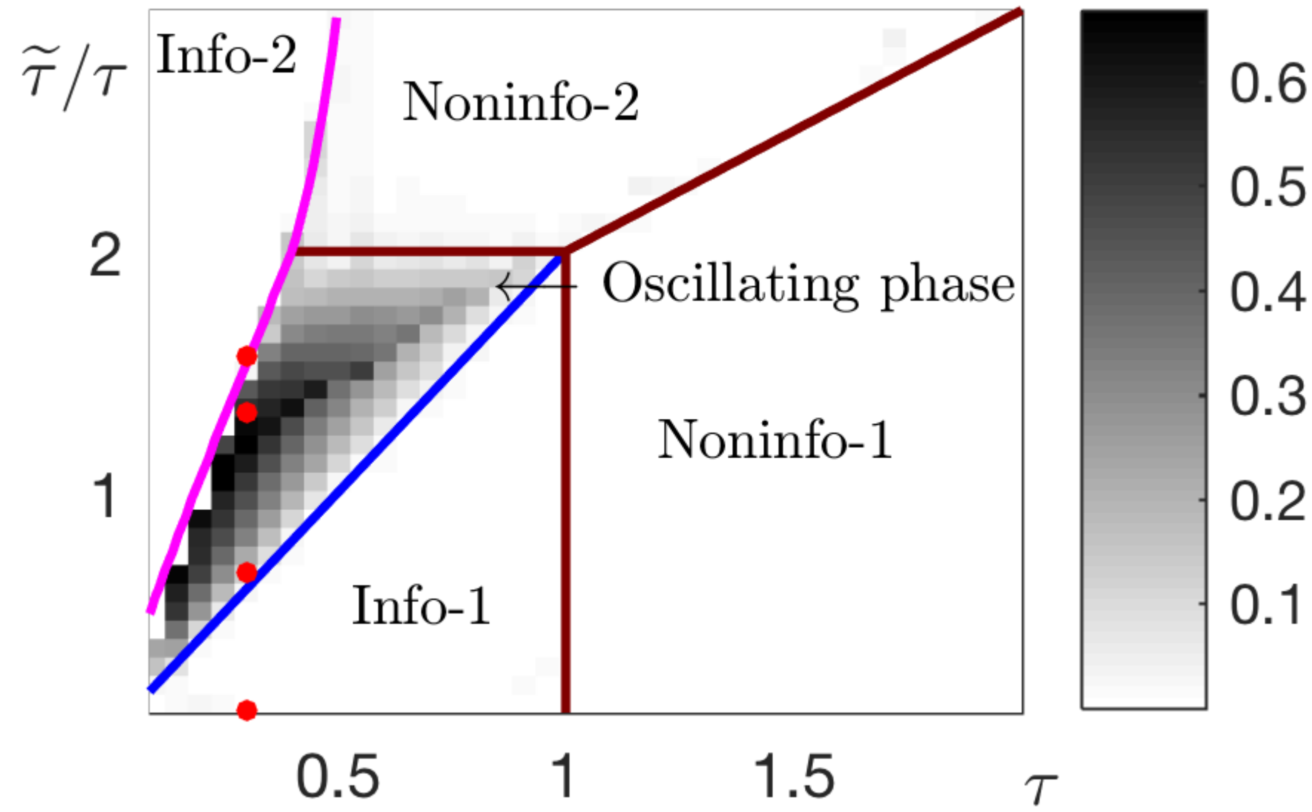




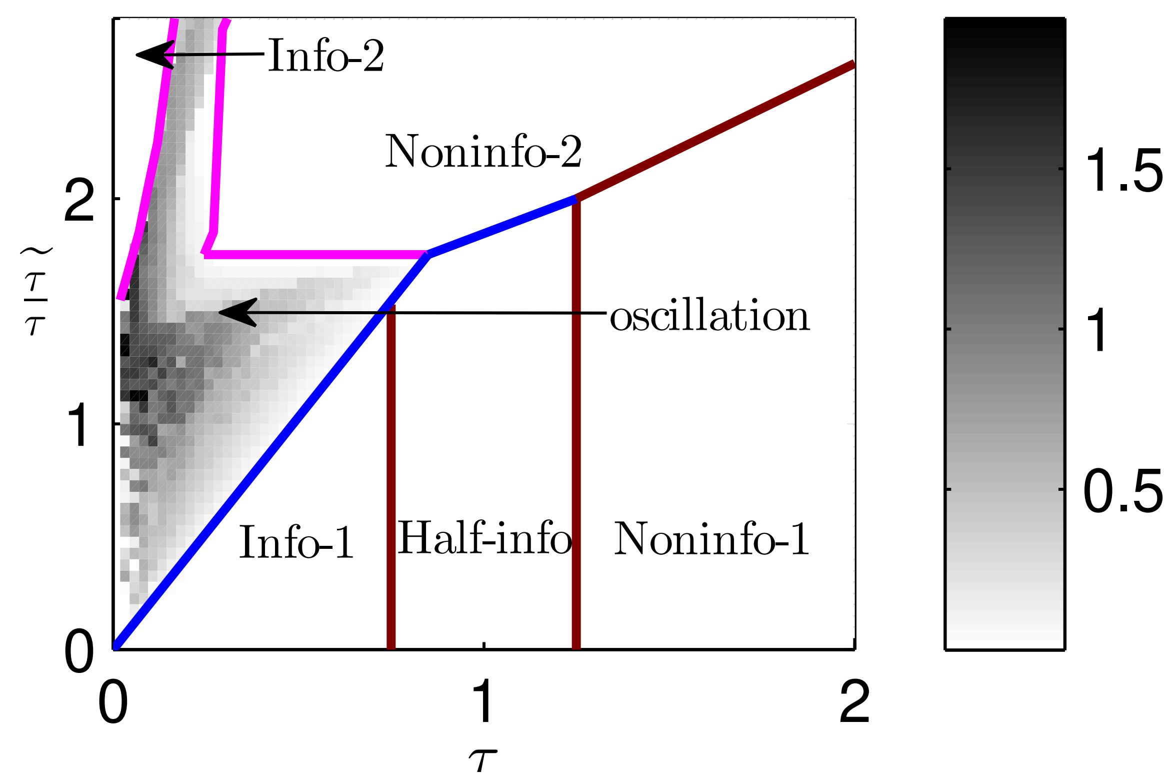
Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Phase diagram

d=1



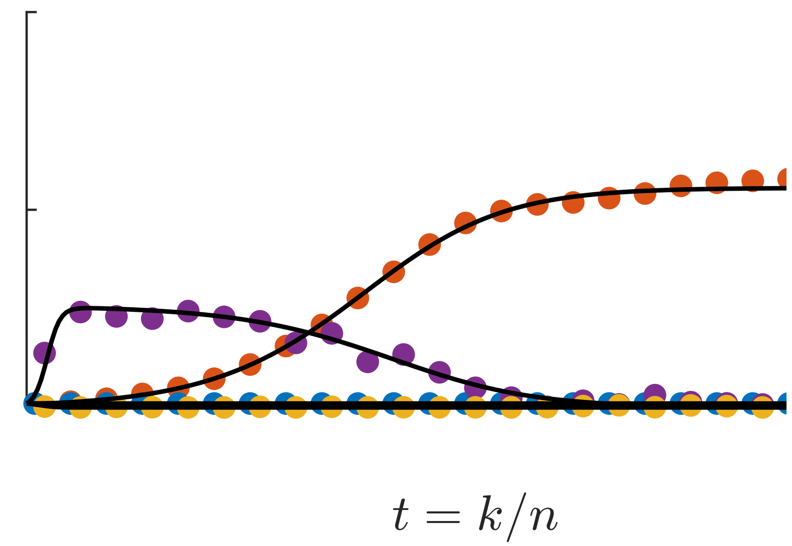
d=2



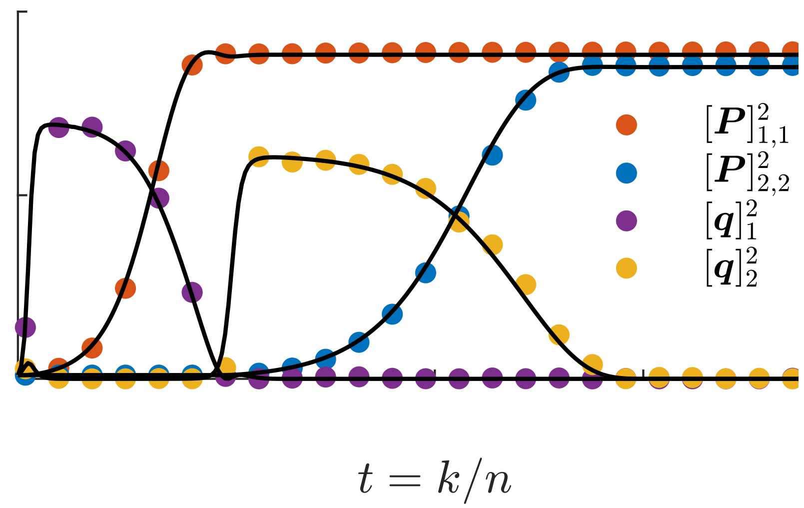
 Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Multi-modes cases

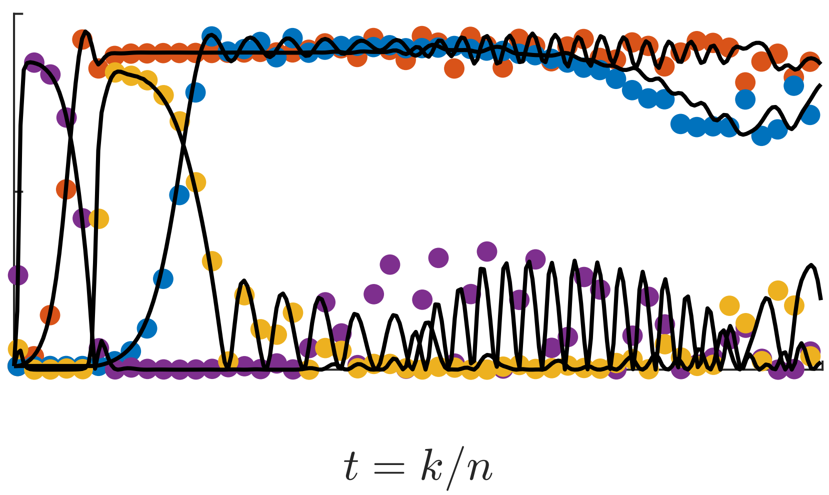
Increase noises



$\sigma^2_{\text{noise}} = 4$

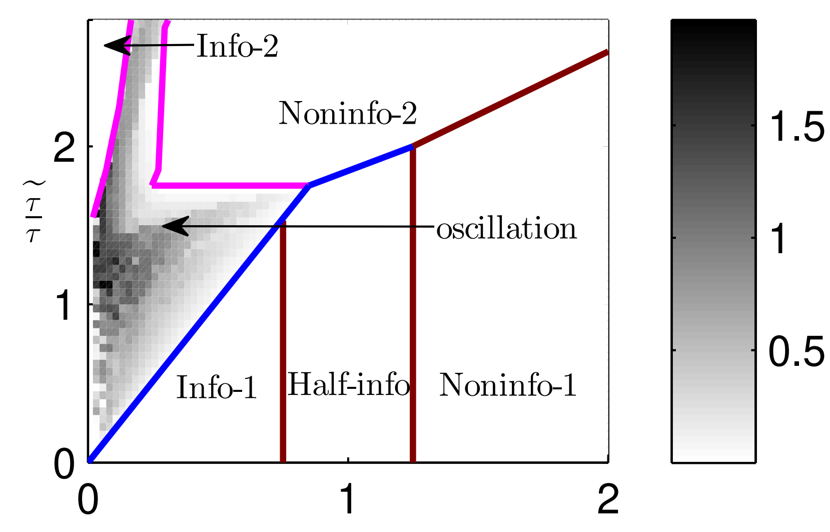


$\sigma^2_{\text{noise}} = 2$



$\sigma^2_{\text{noise}} = 1$ **Noises help converge!**

$$\frac{\tau^{\mathrm{G}}}{2\tau^{\mathrm{D}}} \cdot \Lambda^{\text{signal}}_{\max} \le \tau^{\mathrm{D}} \cdot \sigma^2_{\text{noise}} \le \Lambda^{\text{signal}}_{\min}$$



Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Conclusion

- We present an exact and tractable analysis of the training dynamics of a shallow GAN in high dimensions.

- We analyze the training process at two levels:
  Macroscopic dynamics are deterministic described by a coupled ODE

  Microscopic dynamics are stochastic: The evolution of the detailed weights remains stochastic and it is characterized by an SDE.

- We show that the noise level is essential to the convergence:

  Strong noise leads to failure of feature recovery.

  Weak noise causes oscillation.

 Chuang Wang, Hong Hu, Yue M. Lu, A Solvable High-Dimensional Model of GAN, NeurIPS, 2019

Thanks!

Chuang Wang
wangchuang@ia.ac.cn