Diffusion models learn distributions generated by complex Langevin dynamics

Diaa E. Habibi, Gert Aarts, Lingxiao Wang, Kai Zhou

[1,2]Swansea University — AIMLAC-CDT
[3]RIKEN - iTHEMS
[4]CUHK - Shenzhen
[4]Frankfurt Institute for Advanced Studies

July 29, 2024

Outline

- Motivation
- Complex Langevin
- Diffusion Models
- Results
  - Gaussian case $S_g = \frac{1}{2}\sigma x^2$
  - Quartic case $S_q = S_g + \frac{1}{4}\lambda x^4$
- Summary / Outlook

Motivation

- Systems with a complex Boltzmann weight face the **sign problem**
  - Cannot use conventional MC methods
  - Example: QCD at non-zero baryon density
- **Complex Langevin** (CL) dynamics extend system's degrees of freedom and explore a complexified space via a stochastic process.
  - Extension of Stochastic Quantisation Parisi & Wu 1981
  - Can be successful with models suffering from severe sign problem but can also fail with simple models
  - A lot of work done on testing correctness but issues remain
  - Distributions sampled by CL remain elusive
- **Diffusion Models** (DMs), a machine learning method, have been successful in learning distributions from data.
- Their combination has potential to deepen understanding of CL dynamics.

Complex Langevin

- Partition function $Z = \int dx\, e^{-S(x)}, \qquad S(x) \in \mathbb{C}$
- Complex Langevin dynamics complexify d.o.f :
  $S(x) \mapsto S(z) = S(x + iy)$

$$\dot{x} = K_x + \eta, \qquad K_x = -\mathrm{Re}\partial_z S$$
$$\dot{y} = K_y, \qquad K_y = -\mathrm{Im}\partial_z S$$

- Gaussian noise

$$\langle \eta(t) \rangle = 0, \quad \langle \eta(t)\eta(t') \rangle = 2\delta(t - t')$$

- Complex Langevin for $x(t), y(t) \leftrightarrow$ Fokker-Planck for $P(x, y; t)$

$$\partial_t P(x, y; t) = [\partial_x(\partial_x - K_x) - \partial_y K_y] P(x, y; t)$$

- Generic solution not guaranteed

Complex Langevin

- Complex weight $\rho(x) \leftrightarrow$ Real & positive weight $P(x, y)$
- Considerable effort invested into criteria for correctness to justify results

  Aarts, Seiler, Stamatescu 0912.3360, Aarts, James, Seiler, Stamatescu 1101.3270

$$\int dx\, \rho(x,t) O(x) = \int dxdy\, P(x,y;t) O(x+iy),$$

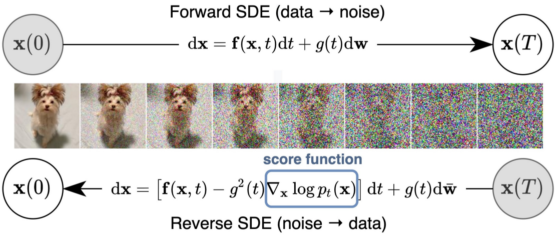
- Holds for holomorphic observables $O(x + iy)$
- If equilibrium $P(x, y)$ is known analytically,

$$\rho(x) = \int dy\, P(x - iy, y)$$

  but hard to verify numerically

Diffusion models

- Diffusion models are a class of generative AI which learn distributions from data
  - Popularly found in *DALL-E & Stable Diffusion*
  - Relies on a stochastic process to learn the score function
  - Relation between DMs and Stochastic Quantisation explored
    Wang, Aarts, Zhou 2309.17082



Source: Song et al. 2011.13456

Diffusion models

Estimate the score of a distribution based on some sample dataset of the target distribution with score-matching

- Approximate score function with a time-dependent score-based model $s_\theta(x, t)$
- Train using the Fisher objective Hyvärinen 2005

$$\mathcal{L}_\theta \propto \mathbb{E}_{p_t(x)} \left[||s_\theta(x, t) - \nabla \log p_t(x)||^2 \right]$$

- Approximate $\nabla \log p_t(x)$ with the score of transition kernel, $\nabla \log p_t(x_t | x_0)$
- In the case of an affine drift $f(x, t)$, the transition kernel is always a Gaussian distribution kernel
- Obtain samples from the target distribution using the estimated score, $s_\theta(x, t) \approx \nabla \log p_t(x)$, via reverse SDE

$$dx = [f(x, t) - g^2(t)\underline{s_\theta(x, t)}]dt + g(t)d\bar{\mathbf{w}}$$

Applications: Two simple examples

Gaussian case

Gaussian action with complex mass parameter

$$S(x) = \frac{1}{2}\sigma x^2, \qquad \sigma = A + iB. \tag{1}$$

This model can be analytically solved!

- CL equation

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} -A & B \\ -B & A \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \eta \\ 0 \end{pmatrix} \tag{2}$$

- Stationary solution of FPE:

$$P(x, y) = N \exp(-\alpha x^2 - \beta y^2 - 2\gamma xy) \tag{3}$$

where $\alpha = A$, $\beta = A\left(1 + \frac{2A^2}{B^2}\right)$ and $\gamma = \frac{A^2}{B}$.

Quartic case

Exact results obtained by

$$Z = \int dx\, e^{-\frac{1}{2}\sigma x^2 - \frac{1}{4}\lambda x^4} = \sqrt{\frac{4\xi}{\sigma}} e^{\xi} K_{-\frac{1}{4}}(\xi), \qquad \sigma = A + iB \quad (4)$$

where $\xi = \sigma^2/(8\lambda)$ and $K_p(\xi)$ is the modified Bessel function of the second kind.

- Moments $\langle x^n \rangle$ are obtained by differentiating w.r.t. σ
- Odd moments vanish
- CL process is contained in a strip $-y_- < y < y_-$, provided $B^2 < 3A^2$ Aarts, Giudice, Seiler 1306:3075

$$y_-^2 = \frac{A}{2\lambda}\left(1 - \sqrt{1 - \frac{B^2}{3A^2}}\right) \quad (5)$$

- But distribution $P(x, y)$ from FPE is not analytically known.

Setup

- Variance Exploding SDE

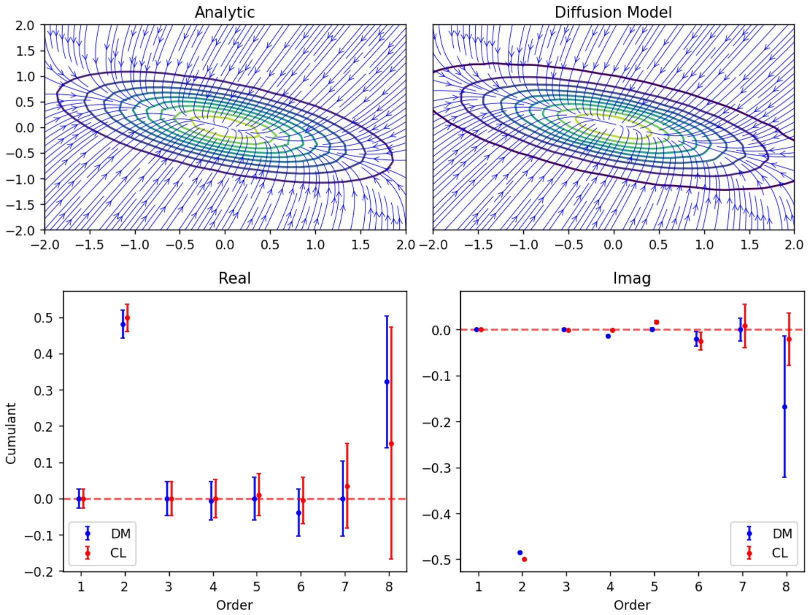
$$dx = \alpha^t d\mathbf{w}$$

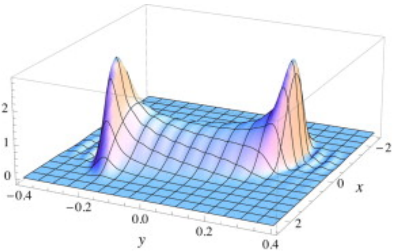
- Score-based model $s_\theta(x, t)$
  - Time-conditioned fully connected NN
  - Hidden layers: [64, 64, 64]
  - Time embedding dims: 128
  - Activation function: LeakyReLU
- Weighted training objective Song et al. 2011.13456

$$\mathcal{L}(\theta, \lambda) := \frac{1}{2} \int_0^T \mathbb{E}_{p_t(x)} \left[\lambda(t) || s_\theta(x, t) - \nabla \log p_t(x) ||_2^2 \right] \, dt$$

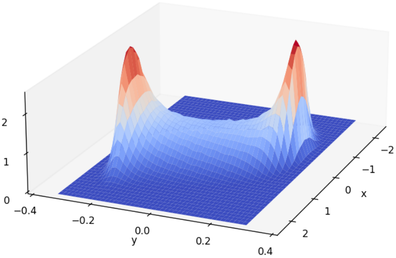
We choose $\lambda(t)$ to be the standard deviation of the noise at time t. For $g(t) = \alpha^t$, $\lambda(t) = \sqrt{\frac{\alpha^{2t} - 1}{2 \log \alpha}}$.

Gaussian case





Source: Aarts, Giudice, Seiler 1306:3075



DM obtained distribution

Table: Results of the observable $\langle O_n(z) \rangle = \langle z^n \rangle / n$ for $\sigma = 1 + i$, $\lambda = 1$.

n	2		4		6		8	
	re	-im	re	-im	re	-im	re	-im
DM	0.2134(3)	0.0745(1)	0.1053(3)	0.06999(13)	0.0964(5)	0.0975(3)	0.1209(11)	0.1741(9)
CL	0.2138(3)	0.0739(1)	0.1059(3)	0.07006(14)	0.0967(4)	0.0981(3)	0.1189(9)	0.1745(9)
Exact	0.214071	0.074005	0.105962	0.070033	0.096741	0.097958	0.118881	0.174170

Quartic case

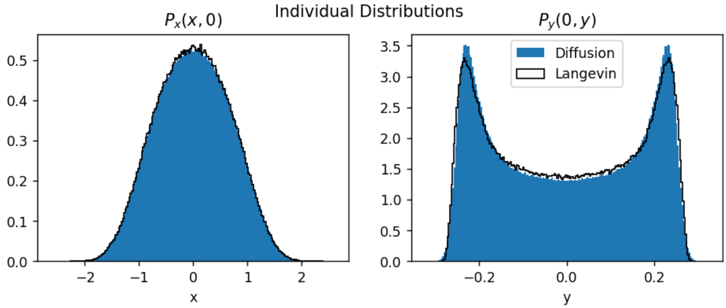


Table: Results of the observable $\langle O_n(z)\rangle = \langle z^n\rangle / n$ for $\sigma = 1 + i$, $\lambda = 1$.

n	2		4		6		8	
	re	-im	re	-im	re	-im	re	-im
DM	0.2134(3)	0.0745(1)	0.1053(3)	0.06999(13)	0.0964(5)	0.0975(3)	0.1209(11)	0.1741(9)
CL	0.2138(3)	0.0739(1)	0.1059(3)	0.07006(14)	0.0967(4)	0.0981(3)	0.1189(9)	0.1745(9)
Exact	0.214071	0.074005	0.105962	0.070033	0.096741	0.097958	0.118881	0.174170

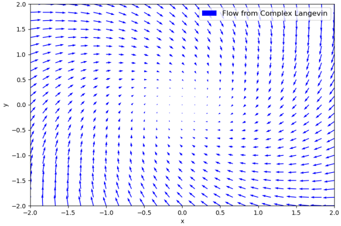
Note

The model doesn't learn the drift of the theory, but instead an effective description of it on the complexified space

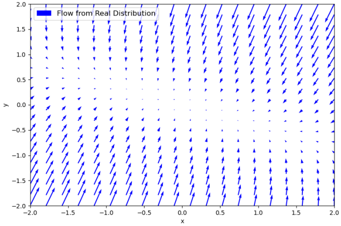
$$\nabla \log P_\theta(x, y) \neq -\partial_z S(z)$$

We can integrate for $P_\theta(x, y)$ but cannot integrate the CL drift since

$$-\partial_y \mathrm{Re}\partial_z S \neq -\partial_x \mathrm{Im}\partial_z S$$



Complex weight flow



Real distribution flow

Summary/Outlook

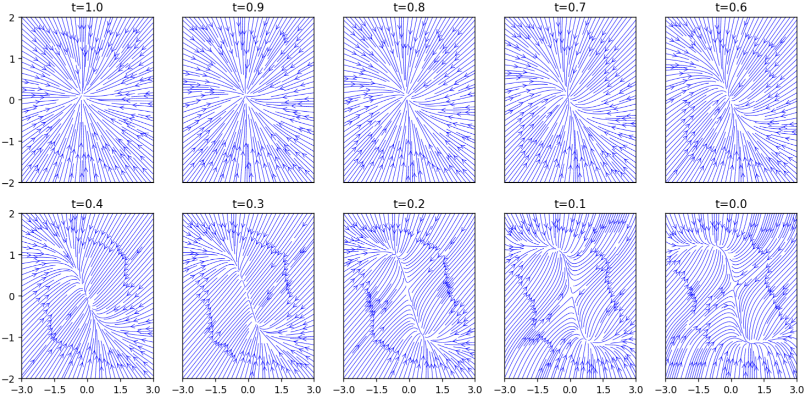
- Diffusion models obtain samples from target distribution
  - Models can't be better than the data they train on
  - Not a cure but we can supplement data using DMs
- Estimated score $s_\theta(x, t) \approx \nabla \log p_t(x)$ lets us analyse properties of the theory
- Extend to lattice theories with sign problem
  - Applying DMs on the lattice has already been explored Wang, Aarts, Zhou 2309.17082
  - Explore conditioning DM on external parameters e.g. chemical potential
- DMs have connections to the renormalisation group flow Colter, Rezchikov 2308.12355

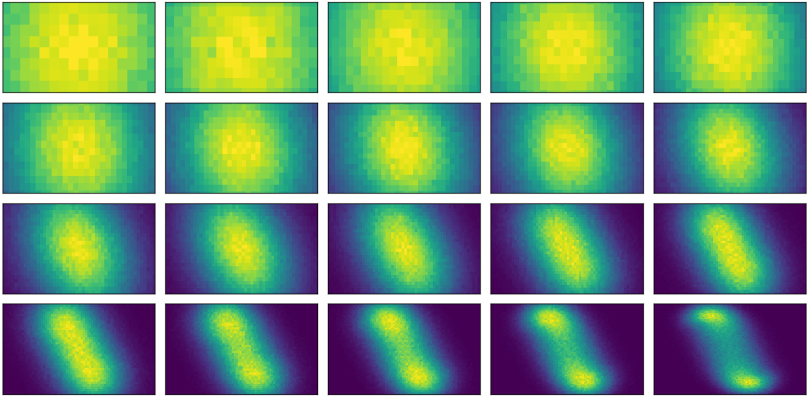
Summary/Outlook

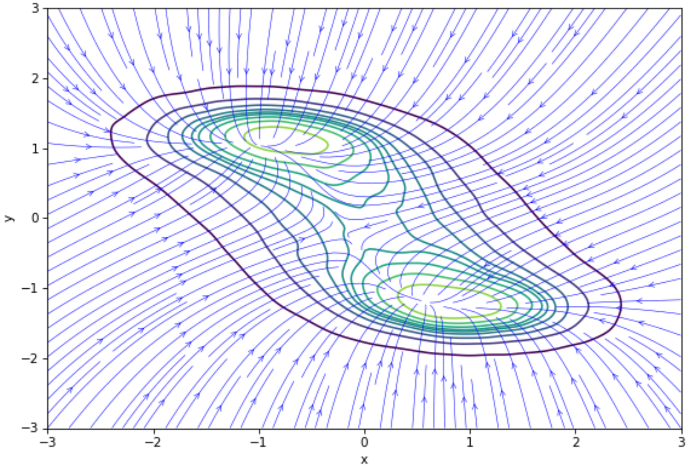
- Diffusion models obtain samples from target distribution
  - Models can't be better than the data they train on
  - Not a cure but we can supplement data using DMs
- Estimated score $s_\theta(x, t) \approx \nabla \log p_t(x)$ lets us analyse properties of the theory
- Extend to lattice theories with sign problem
  - Applying DMs on the lattice has already been explored Wang, Aarts, Zhou 2309.17082
  - Explore conditioning DM on external parameters e.g. chemical potential
- DMs have connections to the renormalisation group flow Colter, Rezchikov 2308.12355

Thank you for listening!

Backup slides


t=1.0
t=0.9
t=0.8
t=0.7
t=0.6
t=0.4
t=0.3
t=0.2
t=0.1
t=0.0
2
1
0
−1
−2
−3.0 −1.5 0.0 1.5 3.0





3
2
1
0
−1
−2
−3
y
−3
−2
−1
0
1
2
3
x

Table: Results of the observable $\langle O_n(z) \rangle = \langle z^n \rangle$ for $\sigma = 1 + i$.

n	2		4		6		8	
	re	-im	re	-im	re	-im	re	-im
DM	0.497(1)	0.491(1)	0.021(1)	1.476(7)	-3.645(26)	3.78(4)	-26.28(11)	0.81(68)
CL	0.4986(7)	0.4990(7)	-0.0018(1)	1.494(5)	-3.748(24)	3.750(26)	-26.42(30)	0.20(3)
Exact	1/2	1/2	0	3/2	-3.75	3.75	-26.25	0