Diffusion models for complex Langevin dynamics

Gert Aarts

with Lingxiao Wang, Kai Zhou and **Diaa Habibi**

JHEP 05 (2024) 060 [2309.17082 [hep-lat]]

NeurIPS workshop 2023 "ML and the Physical Sciences" 2311.03578 [hep-lat]

NeurIPS workshop 2024 "ML and the Physical Sciences" 2410.21212 [hep-lat]

Lattice 2024 2412.01919 [hep-lat]

Swansea University
Prifysgol Abertawe

SIGN 2025, Bern, January 2025

Machine learning is the new playground

many concepts in ML are familiar to theoretical and computational physicists

- neural networks, say, are systems with many fluctuating degrees of freedom
- training – or learning – is a minimisation process, typically achieved with stochastic gradient descent (SGD)
- ML parameters are usually contained in matrices

picture of a playground

keywords: stochastic dynamics, random matrix theory,
non-equilibrium evolution, thermalisation, …

Example: Dyson Brownian motion and SGD

- weight matrices are updated using stochastic gradient descent
- stochastic matrix dynamics: random matrix theory
- Coulomb gas and eigenvalue repulsion: implications for training accuracy
- dependence on learning rate (step size) over batch size (size of fluctuations)

GA, Biagio Lucini and **Chanju Park**, PRE 111 (2025) 1, 015303 [2407.16427 [cond-mat.dis-nn]]

+ **Ouraman Hajizadeh**, NeurIPS 2024 workshop "ML and the Physical Sciences" [2411.13512 [cond-mat.dis-nn]]

+ Matteo Favoni, Lattice 2024 [2412.20496 [hep-lat]]

This talk: GenAI using Diffusion Models

https://encord.com/blog/diffusion-models/

Diffusion models

stochastic dynamics to generate images (configurations)

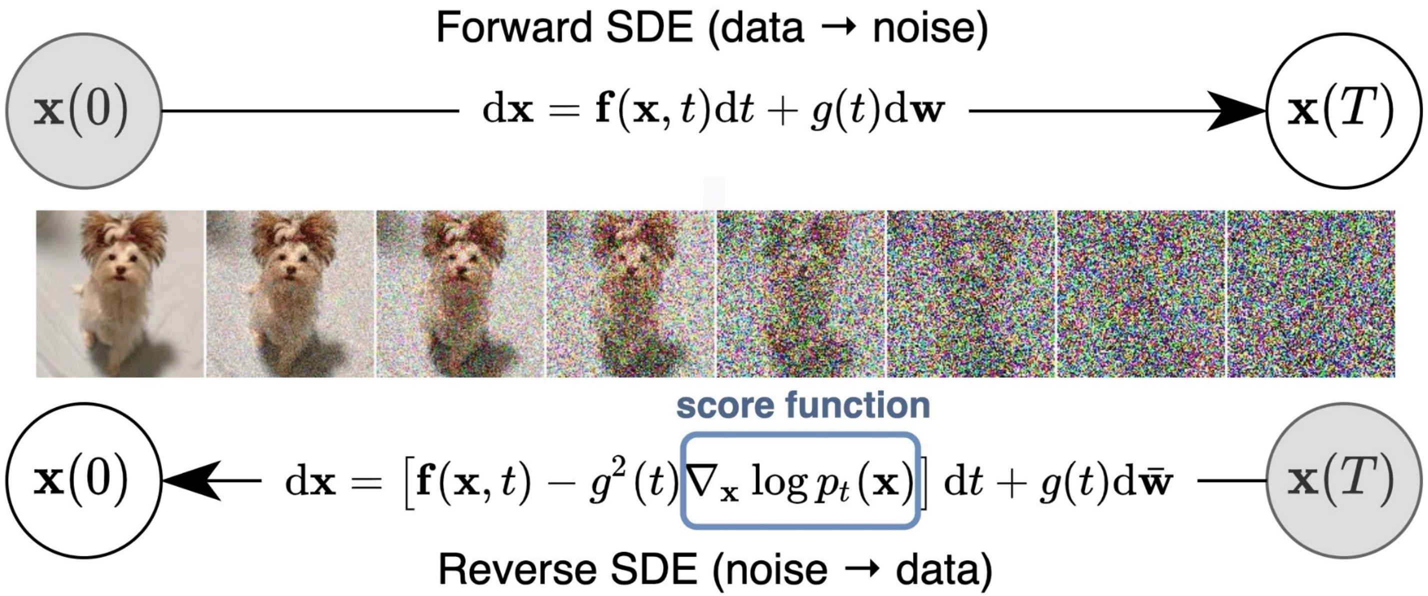
- start with data set of images

- make the images more blurred by applying noise (forward process)

- learn steps in this process

  … and then revert it

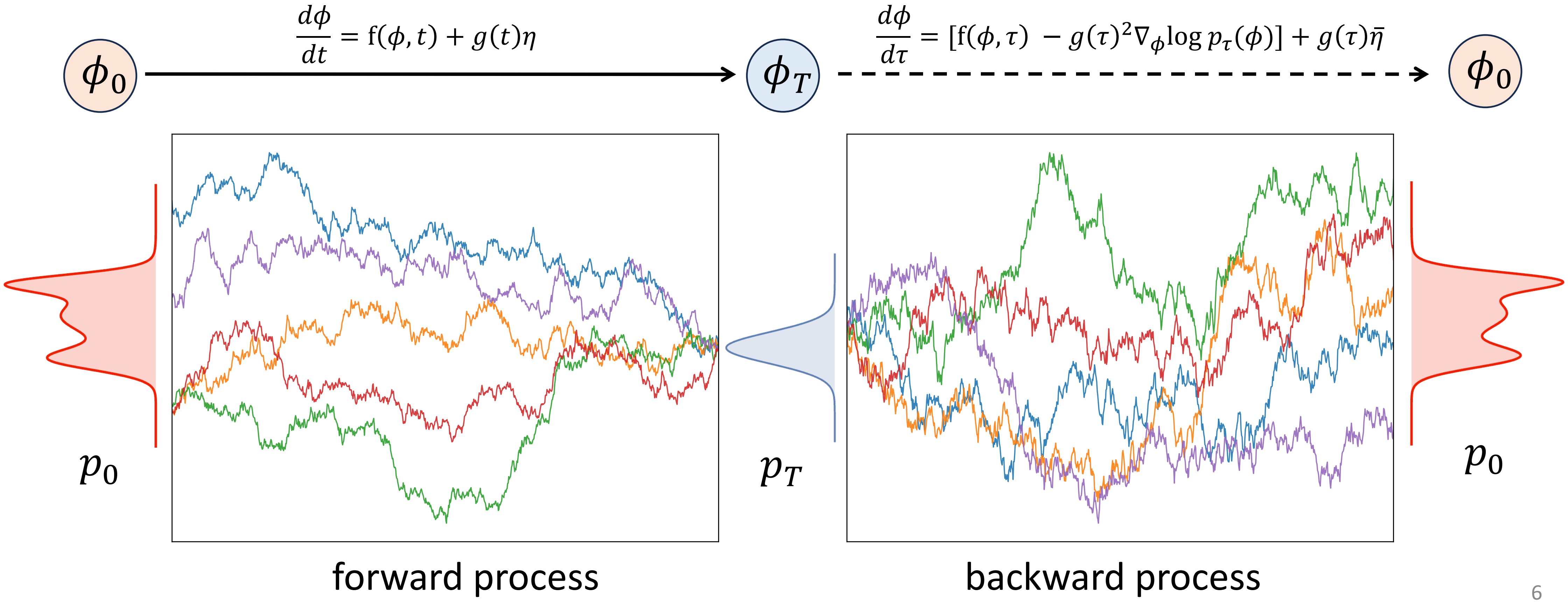
- create new images from noise



https://theaisummer.com/diffusion-models/

Prior and target distributions

- in pictures: p_0 is target (non-trivial), p_T is the prior (easy)



Outline

- some comments on diffusion models and stochastic quantisation
- application in lattice scalar field theory in two dimensions
- correlations: higher n-point functions and cumulants
- application to **sign** and complex action problem: complex Langevin dynamics
- summary and outlook

Diffusion models and stochastic quantisation

- images/configurations are generated during backward process

- stochastic process with time-dependent drift and noise strength

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = g^2(\tau)\nabla_\phi \log P(\phi;\tau) + g(\tau)\eta(x,\tau)$$

- write $P(\phi;\tau) = \frac{e^{-S(\phi,\tau)}}{Z}$ such that $\nabla_\phi \log P(\phi,\tau) = -\nabla_\phi S(\phi,\tau)$

- then $$\frac{\partial \phi(x,\tau)}{\partial \tau} = -g^2(\tau)\nabla_\phi S(\phi,\tau) + g(\tau)\eta(x,\tau)$$

Diffusion models and stochastic quantisation

- then $$\frac{\partial\phi(x,\tau)}{\partial\tau} = -g^2(\tau)\nabla_\phi S(\phi,\tau) + g(\tau)\eta(x,\tau)$$
- very familiar to (lattice) field theorists
- stochastic quantisation (Parisi & Wu 1980)
- path integral quantisation via a stochastic process in fictitious time $$\frac{\partial\phi(x,\tau)}{\partial\tau} = -\nabla_\phi S(\phi) + \sqrt{2}\eta(x,\tau)$$
- stationary solution of associated Fokker-Planck equation $P(\phi) \sim e^{-S(\phi)}$

Diffusion models and stochastic quantisation

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = g^2(\tau)\nabla_\phi \log P(\phi;\tau) + g(\tau)\eta(x,\tau)$$

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = -\nabla_\phi S(\phi) + \sqrt{2}\eta(x,\tau)$$

similarities and differences:

✓ SQ: fixed drift, determined from known action

constant noise variance (but can be generalised using kernels)

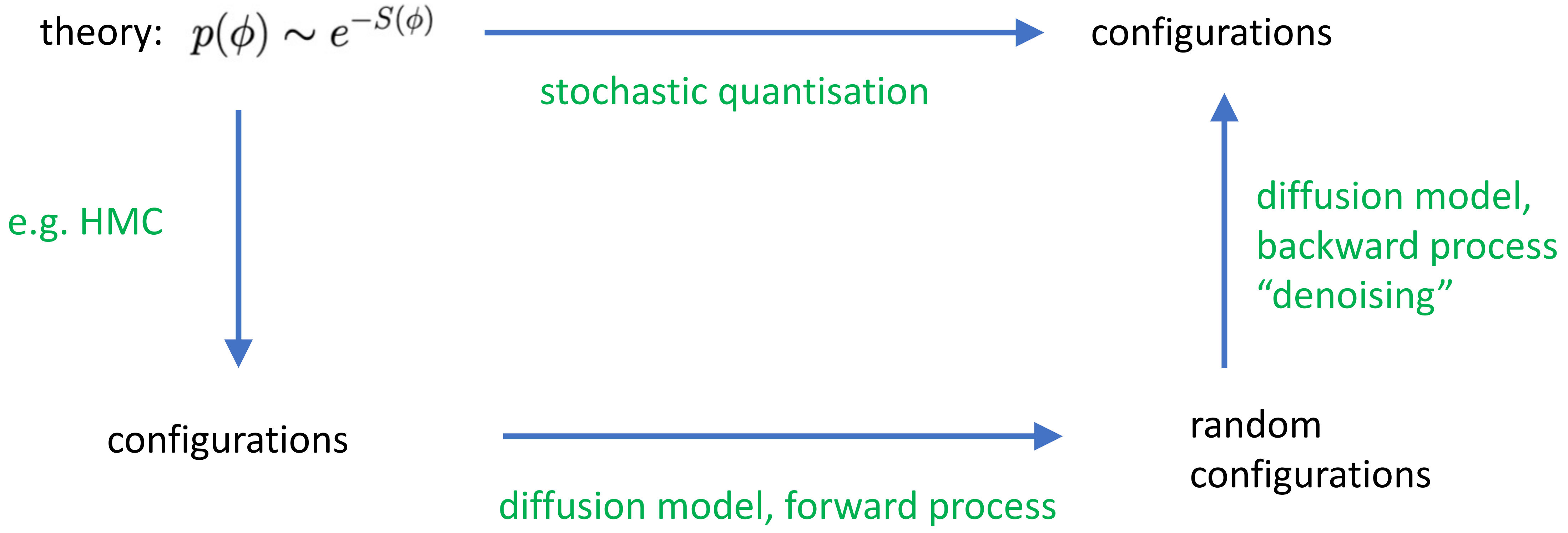
thermalisation followed by long-term evolution in equilibrium

✓ DM: drift and noise variance time-dependent, learn from data

evolution between $0 \leq \tau \leq T = 1$, many short runs

Diffusion models and stochastic quantisation

- diffusion models as an alternative approach to stochastic quantisation

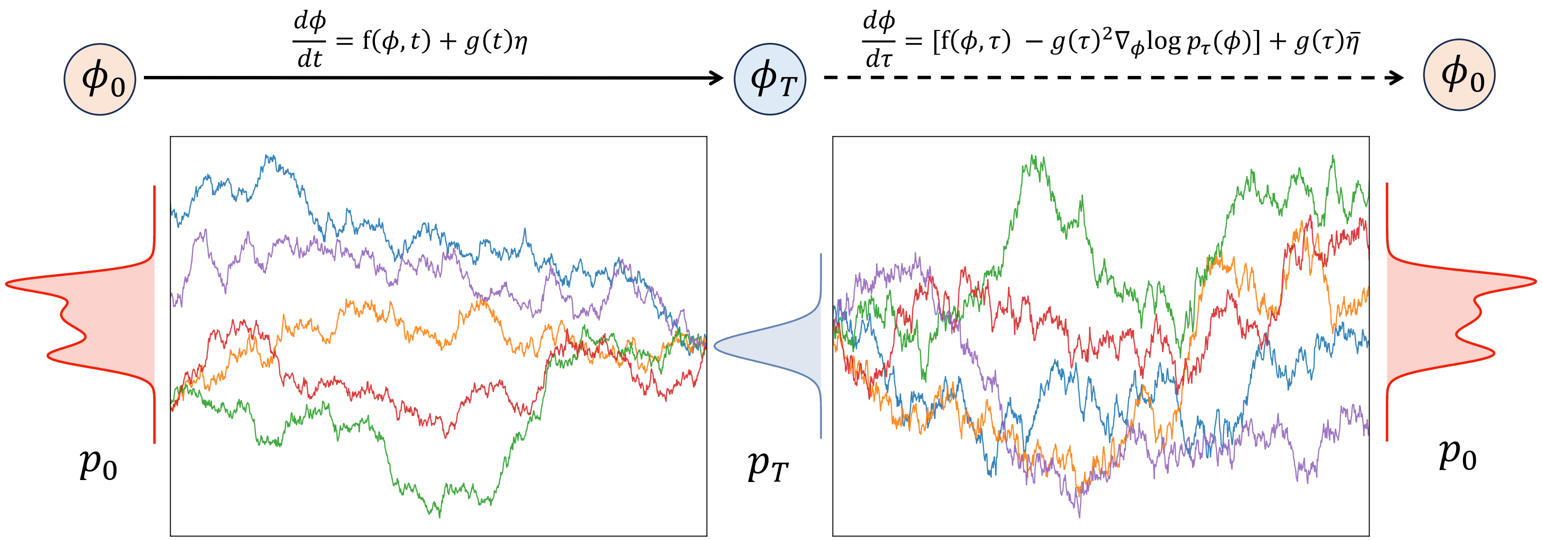


Score matching: learn the drift for backward process

- one degree of freedom, variance-expanding scheme: $\dot{x}(t) = g(t)\eta(t) \qquad \eta \sim \mathcal{N}(0,1)$

- time-dependent distribution $P(x,t) = P_t(x)$ describes forward and backward process

- so-called **score** $\nabla \log P_t(x)$ is not known, needs to be "learnt" during forward process



Score matching: learn the drift for backward process

- one degree of freedom, variance-expanding scheme: $\dot{x}(t) = g(t)\eta(t) \qquad \eta \sim \mathcal{N}(0,1)$
- time-dependent distribution $P(x,t) = P_t(x)$ describes forward and backward process
- so-called **score** $\nabla \log P_t(x)$ is not known, needs to be "learnt"
- loss function $\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x)}\left[\sigma^2(t)\, \| s_\theta(x,t) - \nabla \log P_t(x) \|^2\right] \qquad \sigma^2(t) = \int_0^t ds\, g^2(s)$
- $s_\theta(x,t)$ approximates score, vector field learnt by some neural network
- introduce conditional distribution $P_t(x) = \int dx_0\, P_t(x|x_0) P_0(x_0)$ initial data $P_0(x_0)$

$$P_t(x) = \int dx_0\, P_t(x|x_0) P_0(x_0)$$

Score matching: learn the drift

- loss function $\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x)}\left[\sigma^2(t)\, \|s_\theta(x,t) - \nabla \log P_t(x)\|^2\right]$
- diffusion process $\dot{x}(t) = g(t)\eta(t)$ easily solved $x(t) = x_0 + \sigma(t)\eta(t)$ $\qquad \sigma^2(t) = \int_0^t ds\, g^2(s)$
- conditional distribution $P_t(x|x_0) = \mathcal{N}(x; x_0, \sigma^2(t)) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} e^{-(x-x_0)^2/(2\sigma^2(t))}$
- and hence $\nabla \log P_t(x_t|x_0) = -(x_t - x_0)/\sigma^2(t)$
- loss function $\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x_t)}\left[\left\|\sigma(t) s_\theta(x_t,t) + \frac{x_t - x_0}{\sigma(t)}\right\|^2\right]$

$$= \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x_t)}\left[\|\sigma(t) s_\theta(x_t,t) + \eta(t)\|^2\right]$$

tractable, computable

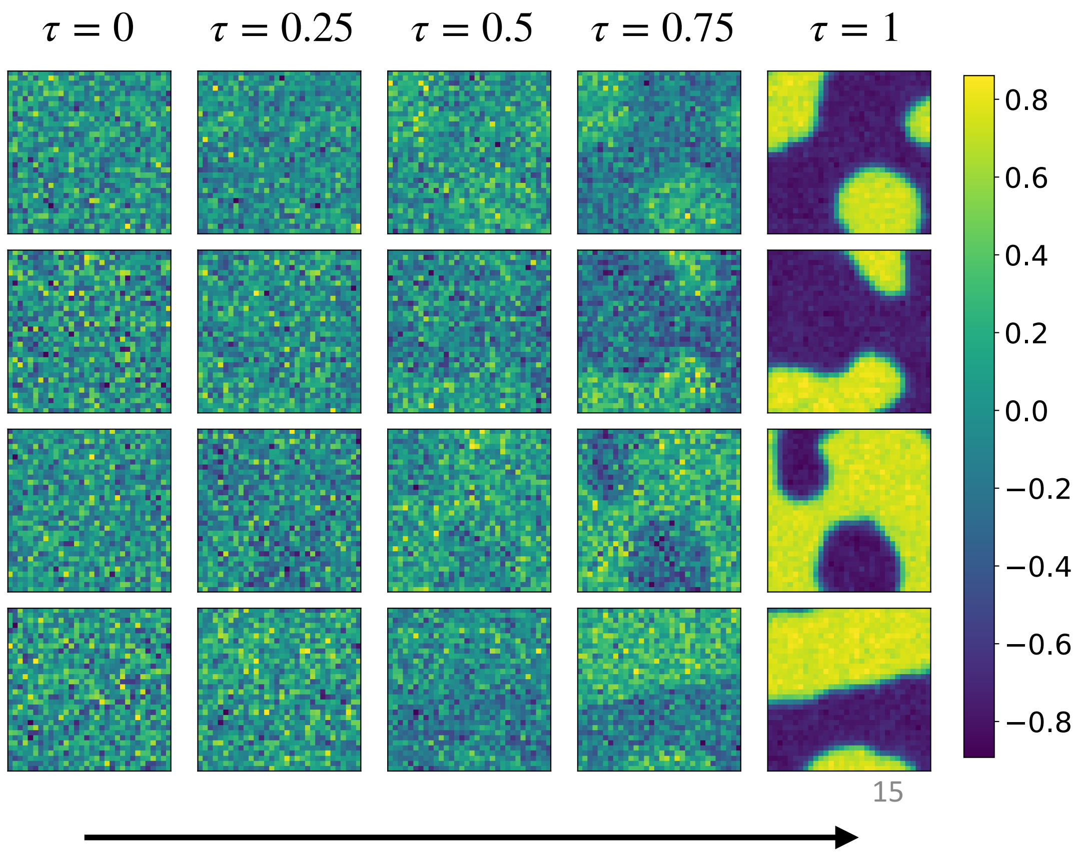
Diffusion model for 2d ϕ^4 scalar theory

- 32^2 lattice, choice of action parameters in symmetric and broken phase
- training data set generated using Hybrid Monte Carlo (HMC)

- variance expanding DM trained using U-Net architecture

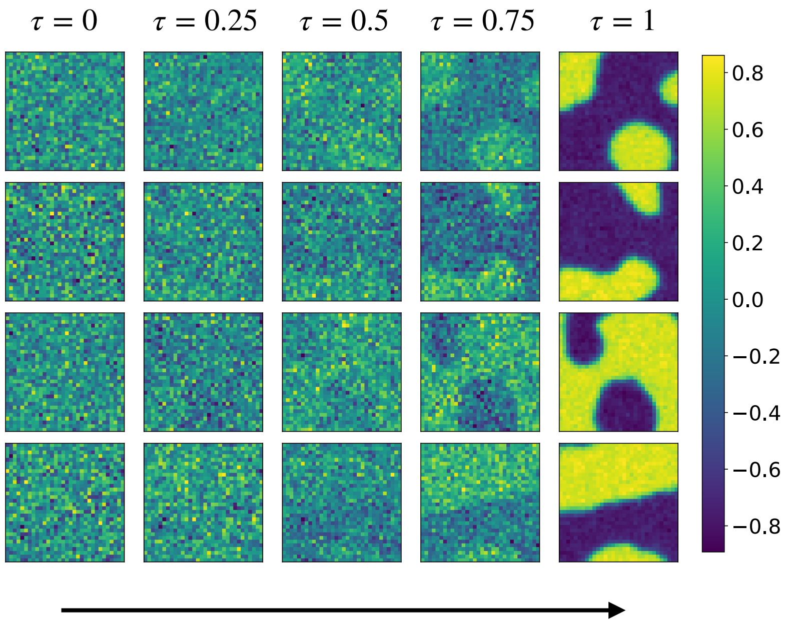
generating configurations:

- broken phase
- "denoising" (backward process)
- large-scale clusters emerge, as expected



L Wang, GA, K Zhou, JHEP 05 (2024) 060 [2309.17082 [hep-lat]]

Diffusion models



ok, so it seems to work: many questions

- correlations: how are they destroyed and rebuilt?
- usually attention is on two-point function or variance
- higher n-point functions contain interactions in field theory
- essential for applications in field theory, correlations = interactions
- focus on moments and cumulants

discuss forward and backward process in more detail

GA, D Habibi, L Wang, K Zhou [2410.21212 [hep-lat]]

Diffusion models in more detail

- forward process $\dot{x}(t) = K(x(t), t) + g(t)\eta(t)$ $0 \leq t \leq T$

  noise profile $g(t) = \sigma^{t/T}$

- backward process

$$x'(\tau) = -K(x(\tau), T-\tau) + g^2(T-\tau)\partial_x \log P(x, T-\tau) + g(T-\tau)\eta(\tau)$$

score

$\tau = T - t$

two main schemes

- variance-expanding (VE): no drift $K(x,t) = 0$
- variance-preserving (VP) or denoising diffusion probabilistic models (DDPMs):

  linear drift $K(x(t), t) = -\frac{1}{2}k(t)x(t)$

assume first moment vanishes

$$x_0 \to x_0 - \mathbb{E}_{P_0}[x_0]$$

Solve forward process

- forward process $\dot{x}(t) = K(x(t), t) + g(t)\eta(t)$ $\quad K(x(t), t) = -\frac{1}{2}k(t)x(t)$
- initial data from target ensemble $x_0 \sim P_0(x_0)$
- solution $x(t) = x_0 f(t, 0) + \int_0^t ds\, f(t, s)g(s)\eta(s)$ $\quad f(t, s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$
- second moment/cumulant/variance $\kappa_2(t) = \mu_2(t) = \mu_2(0) f^2(t, 0) + \Xi(t)$

$$\Xi(t) = \int_0^t ds \int_0^t ds'\, f(t, s) f(t, s') g(s) g(s') \mathbb{E}_\eta[\eta(s)\eta(s')] = \int_0^t ds\, f^2(t, s) g^2(s)$$

$$f(t,s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$$

Higher-order moments and cumulants

- moments $\mu_n(t) = \mathbb{E}[x^n(t)]$ and cumulants $\kappa_n(t)$: straightforward algebra

$$\kappa_3(t) = \mu_3(t) = \kappa_3(0) f^3(t,0)$$

$$\mu_4(t) = \mu_4(0) f^4(t,0) + 6\mu_2(0) f^2(t,0)\Xi(t) + 3\Xi^2(t)$$

$$\kappa_4(t) = \mu_4(t) - 3\mu_2^2(t) = \left[\mu_4(0) - 3\mu_2^2(0)\right] f^4(t,0) = \kappa_4(0) f^4(t,0)$$

$$\kappa_5(t) = \left[\mu_5(0) - 10\mu_3(0)\mu_2(0)\right] f^5(t,0) = \kappa_5(0) f^5(t,0)$$

→ $$\kappa_{n>2}(t) = \kappa_n(0) f^n(t,0)$$

variance-expanding scheme: no drift

$$f(t,0) = 1$$

higher cumulants conserved!

$$\Xi(t) = \int_0^T ds\, f^2(t,s) g^2(s)$$

Proof to all orders

- generating functionals: average over both noise and target distributions

moments $Z[J] = \mathbb{E}[e^{J(t)x(t)}]$ cumulants $W[J] = \log Z[J]$

- noise average $$Z_\eta[J] = \mathbb{E}_\eta[e^{J(t)x(t)}] = \frac{\int D\eta\, e^{-\frac{1}{2}\int_0^t ds\, \eta^2(s) + J(t)\left[x_0 f(t,0) + \int_0^t ds\, f(t,s) g(s) \eta(s)\right]}}{\int D\eta\, e^{-\frac{1}{2}\int_0^t ds\, \eta^2(s)}}$$

- full average $$Z[J] = \mathbb{E}[e^{J(t)x(t)}] = e^{\frac{1}{2}J^2(t)\Xi(t)} \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$

- cumulant generator $$W[J] = \log Z[J] = \frac{1}{2} J^2(t)\Xi(t) + \log \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$

$$f(t,s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$$

$$\Xi(t) = \int_0^T ds\, f^2(t,s) g^2(s)$$

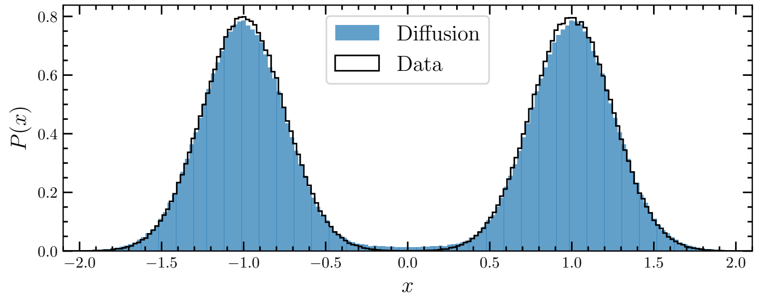
Proof to all orders: cumulants

- cumulant generator $$W[J] = \log Z[J] = \frac{1}{2} J^2(t)\Xi(t) + \log \int dx_0\, P_0(x_0) e^{J(t) x_0 f(t,0)}$$

- 2nd cumulant $$\kappa_2(t) = \frac{d^2 W[J]}{dJ(t)^2}\bigg|_{J=0} = \Xi(t) + \mathbb{E}_{P_0}[x_0^2] f^2(t,0)$$ ✓

- higher-order cumulants $$\kappa_{n>2}(t) = \frac{d^n W[J]}{dJ(t)^n}\bigg|_{J=0} = \frac{d^n}{dJ(t)^n} \log \mathbb{E}_{P_0}[e^{J(t) x_0 f(t,0)}]\bigg|_{J=0} = \kappa_n(0) f^n(t,0)$$ ✓

Toy model: two-peak distribution

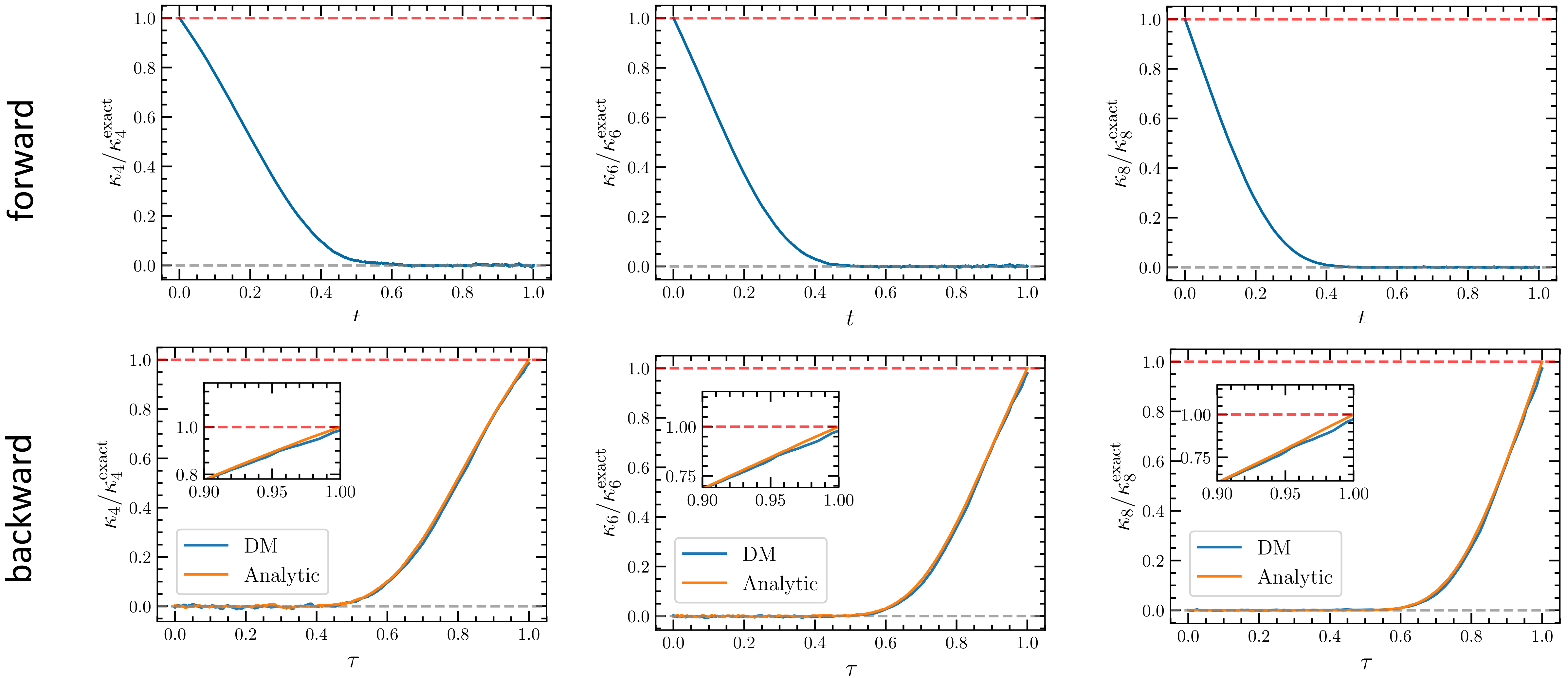


- test the predictions in simple zero-dimensional model
- sum of two Gaussians $P_0(x) = \frac{1}{2}\left[\mathcal{N}(x;\mu_0,\sigma_0^2) + \mathcal{N}(x;-\mu_0,\sigma_0^2)\right]$
- exactly solvable, all even cumulants non-zero, time-dependent score is known analytically
- quickly show higher-order cumulants, see paper for details

$$\kappa_{n>2}(t) = \kappa_n(0) f^n(t,0)$$

$$f(t,0) \to 0$$

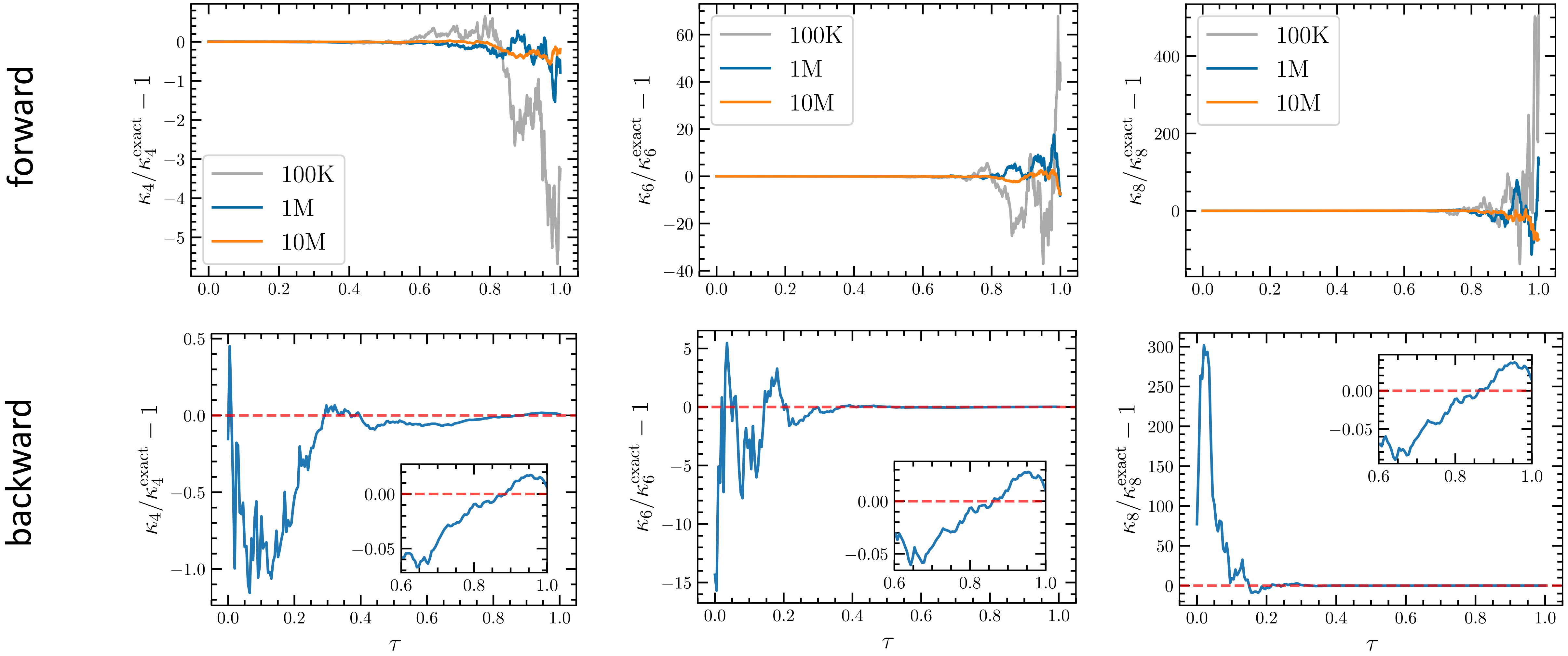
4th, 6th, 8th cumulant with drift (DDPM)



analytic = analytic score

$$\kappa_{n>2}(t) = \kappa_n(0)$$

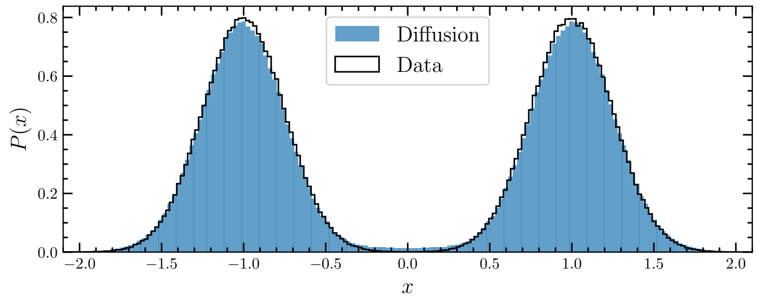
4^{th}, 6^{th}, 8^{th} cumulant without drift



Higher-order cumulants

- with drift (DDPM): cumulants go to zero, distribution becomes normal
- without drift (variance-expanding): higher-order cumulants are conserved, up to numerical cancellations, required between moments which increase in time
- initial conditions for backward process taken from normal distribution
- score has higher-order cumulants encoded: cumulants are reconstructed

Comparison between schemes



	κ_2	κ_4	κ_6	κ_8
Exact	1.0625	−2	16	−272
Data	1.0624(5)	−2.000(2)	16.00(2)	−272.0(6)
Variance expanding	1.0692(6)	−2.001(2)	16.03(3)	−272.7(6)
Variance preserving (DDPM)	1.0609(5)	−1.976(2)	15.72(2)	−265.6(6)

expectation values at the end of the backward process

✓ variance-expanding scheme slightly outperforms variance-preserving scheme

Two-dimensional scalar fields

extension to scalar fields trivial: each lattice point is treated separately

- forward $\partial_t \phi(x,t) = K[\phi(x,t),t] + g(t)\eta(x,t)$
- backward $\partial_\tau \phi(x,\tau) = -K[\phi(x,\tau),T-\tau] + g^2(T-\tau)\nabla_\phi \log P(\phi,T-\tau) + g(T-\tau)\eta(x,\tau)$
- two-point function $G(x,y;t) \equiv \mathbb{E}[\phi(x,t)\phi(y,t)] = \mathbb{E}_{P_0}[\phi_0(x)\phi_0(y)]f^2(t,0) + \Xi(t)\delta(x-y)$
- moments $\mu_n(x,t) = \mathbb{E}[\phi^n(x,t)]$ independent of x

$$\Xi(t) = \int_0^T ds\, f^2(t,s) g^2(s)$$

Generating functionals

full path integral
with sources

- moment generating

$$Z[J] = \mathbb{E}[e^{J(x,t)\phi(x,t)}] = e^{\frac{1}{2}J^2(x,t)\Xi(t)} \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x)f(t,0)}$$

- cumulant generating

$$W[J] = \log Z[J] = \frac{1}{2} J^2(x,t)\Xi(t) + \log \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x)f(t,0)}$$

- higher-order cumulants

$$\kappa_{n>2}(t) = \frac{\delta^n W[J]}{\delta J(x,t)^n}\Big|_{J=0} = \frac{\delta^n}{\delta J(x,t)^n} \log \mathbb{E}_{P_0}[e^{J(x,t)\phi_0(x)f(t,0)}]\Big|_{J=0}$$

variance preserving

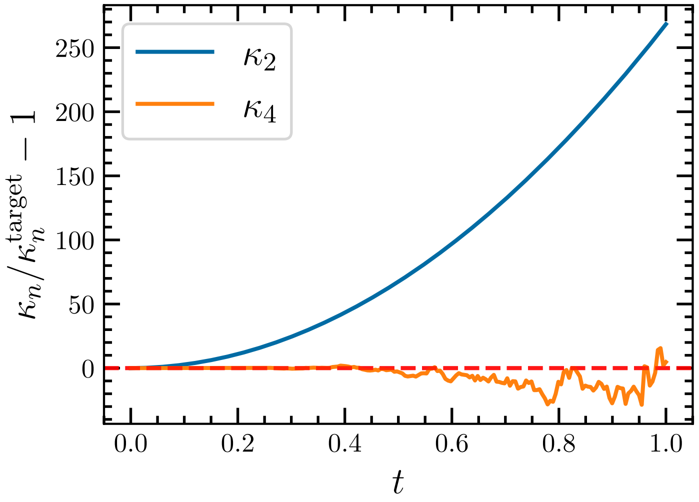
$$f(t,0) \to 0$$

variance expanding

$$f(t,0)) = 1$$

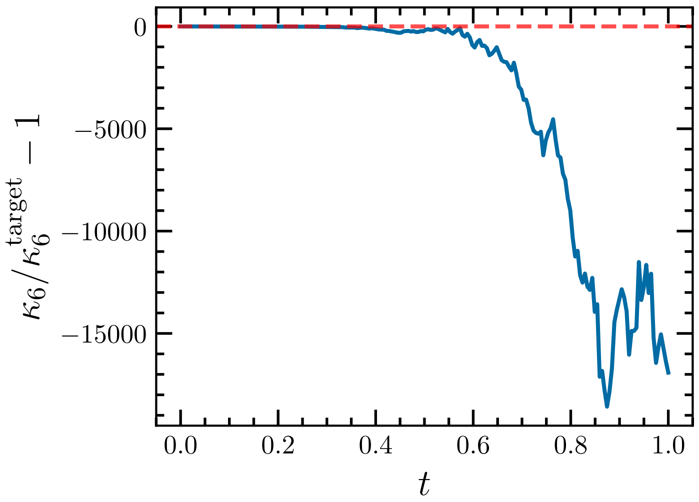
2nd, 4th, 6th cumulant without drift

forward

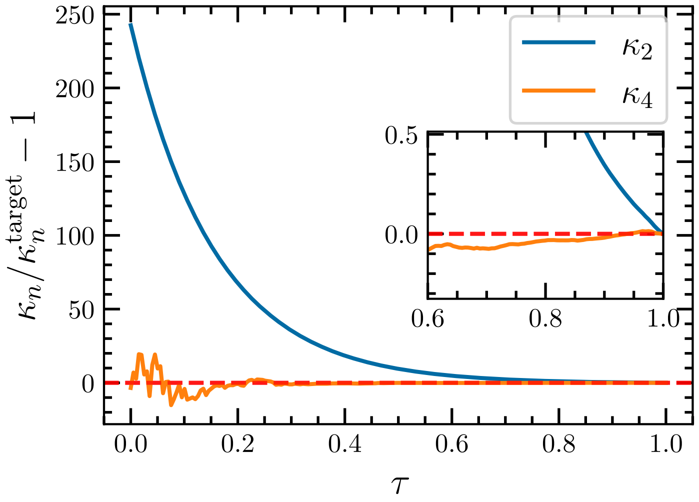


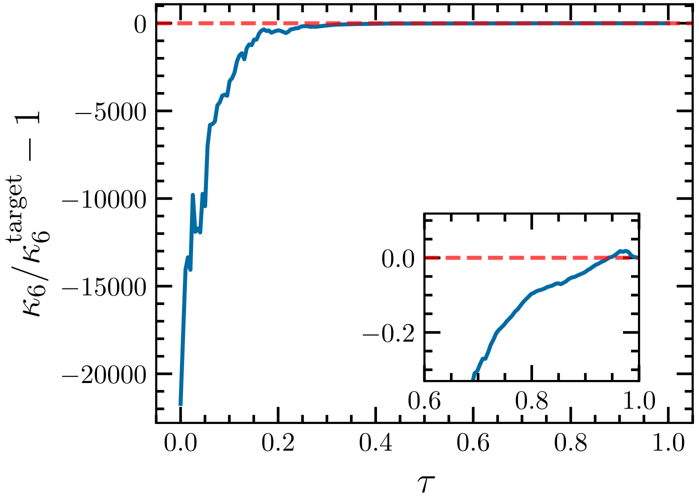
κ_2, κ_4

κ_6



backward





Comparison: trained diffusion model

	κ_2	κ_4	κ_6	κ_8
HMC (normalised)	0.39597(4)	−0.29453(6)	0.90108(28)	−5.8689(25)
Diffusion model	0.39598(4)	−0.29454(7)	0.90113(32)	−5.8694(28)

ϕ^4 theory: $32^2, \kappa = 0.4, \lambda = 0.022, 10^5$ configurations

expectation values at the end of the backward process

excellent agreement

Machine Learning and the Physical Sciences

Workshop at the 38th conference on Neural Information Processing Systems (NeurIPS)

December 15, 2024

https://ml4physicalsciences.github.io/2024/

Best Paper Awards

Sponsored by Apple. Awardees get an iPhone 16 Pro.

BEST 'AI FOR PHYSICS' PAPER AWARD

Robust Emulator for Compressible Navier-Stokes using Equivariant Geometric Convolutions

Wilson G. Gregory, David W Hogg, Kaze W. K. Wong, Soledad Villar

[paper]

BEST 'PHYSICS FOR AI' PAPER AWARD

Higher-order cumulants in diffusion models

Gert Aarts, Diaa Eddin Habibi, Lingxiao Wang, Kai Zhou

[paper] [poster]

Sign problem, complex Langevin dynamics and diffusion models

Diaa Habibi, GA, Lingxiao Wang, Kai Zhou

Lattice 2024 2412.01919 [hep-lat] and in preparation

Stochastic quantisation: complex actions

- stochastic quantisation not limited to real-valued distributions/actions
- extend Langevin process to complex manifold: complex Langevin dynamics (Parisi 1981)

$$z \sim \rho(z) \in \mathbb{C} \quad \Rightarrow \quad x, y \sim P(x, y) \in \mathbb{R}$$

- convergence not guaranteed, no general solution of Fokker-Planck equation
- a posteriori justification (GA, Seiler, Stamatescu 2009, Nagata, Nishimura, Shimasaki 2016)

- many talks at this meeting

(Complex) Langevin dynamics

- observables $\langle O(x)\rangle = \int dx\, \rho(x) O(x)$ $\rho(x) = \frac{1}{Z}\exp[-S(x)]$ $Z = \int dx\, \rho(x)$
- Langevin equation and drift $\dot{x}(t) = K[x(t)] + \eta(t)$ $K(x) = \frac{d}{dx}\log\rho(x) = -\frac{dS(x)}{dx}$
- Fokker-Planck equation (FPE) $\partial_t \rho(x;t) = \partial_x\left[\partial_x - K(x)\right]\rho(x;t)$
- what if weight is complex? drift is complex, FPE only formal
- complexify degrees of freedom $z \to x + iy$

Complex Langevin dynamics

- complexify degrees of freedom $z \to x + iy$
- Langevin equation and drift
$$\dot{z}(t) = K[z(t)] + \eta(t), \qquad K(z) = \frac{d}{dz}\log\rho(z) = -\frac{dS(z)}{dz}$$
- take real and imaginary part $N_x - N_y = 1$
$$\dot{x}(t) = K_x + \eta_x(t), \qquad K_x = \mathrm{Re}\frac{d}{dz}\log\rho(z), \qquad \langle\eta_x(t)\eta_x(t')\rangle = 2N_x\delta(t-t')$$
$$\dot{y}(t) = K_y + \eta_y(t), \qquad K_y = \mathrm{Im}\frac{d}{dz}\log\rho(z), \qquad \langle\eta_y(t)\eta_y(t')\rangle = 2N_y\delta(t-t')$$
- FPE $\partial_t P(x,y;t) = \left[\partial_x\left(N_x\partial_x - K_x\right) + \partial_y\left(N_y\partial_y - K_y\right)\right]P(x,y;t)$ $\qquad P(x,y;t) \geq 0$
- observables $\langle O[x(t)+iy(t)]\rangle_\eta = \int dxdy\, P(x,y;t)O(x+iy)$

Complex Langevin dynamics

- FPE $$\partial_t P(x,y;t) = \left[\partial_x\left(N_x\partial_x - K_x\right) + \partial_y\left(N_y\partial_y - K_y\right)\right]P(x,y;t)$$
- cannot be solved, non-integrable $$\partial_x K_y \neq \partial_y K_x$$
- formal justification $$\int dxdy\, P(x,y)O(x+iy) = \int dx\, \rho(x)O(x)$$
- relation (cannot be verified in practice) $$\rho(x) = \int dy\, P(x-iy,y)$$
- instead, a posteriori criteria for correctness

GA, E Seiler, IO Stamatescu, *Phys. Rev. D* **81** (2010) 054508 [0912.3360]
GA, F James, E Seiler, IO Stamatescu, *Eur. Phys. J. C* **71** (2011) 1756 [1101.3270]

Complex Langevin distributions

- FPE $$\partial_t P(x,y;t) = \left[\partial_x \left(N_x \partial_x - K_x\right) + \partial_y \left(N_y \partial_y - K_y\right)\right] P(x,y;t)$$

real noise:

$$N_x = 1, N_y = 0$$

- want to describe/understand this distribution
  - further sampling
  - criteria for correctness
  - (modify process)

$$P(x,y;t) \geq 0$$

- use diffusion model, learn from CL generated data
- diffusion model does not care what the origin of the data is
- note: no solution to the sign problem if CL fails

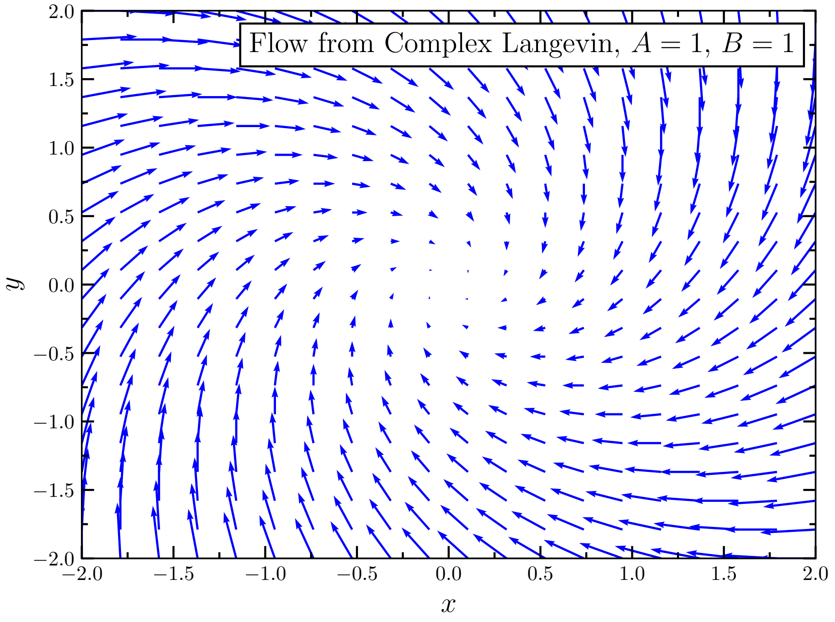
introductory lectures
1512.05145 [hep-lat])

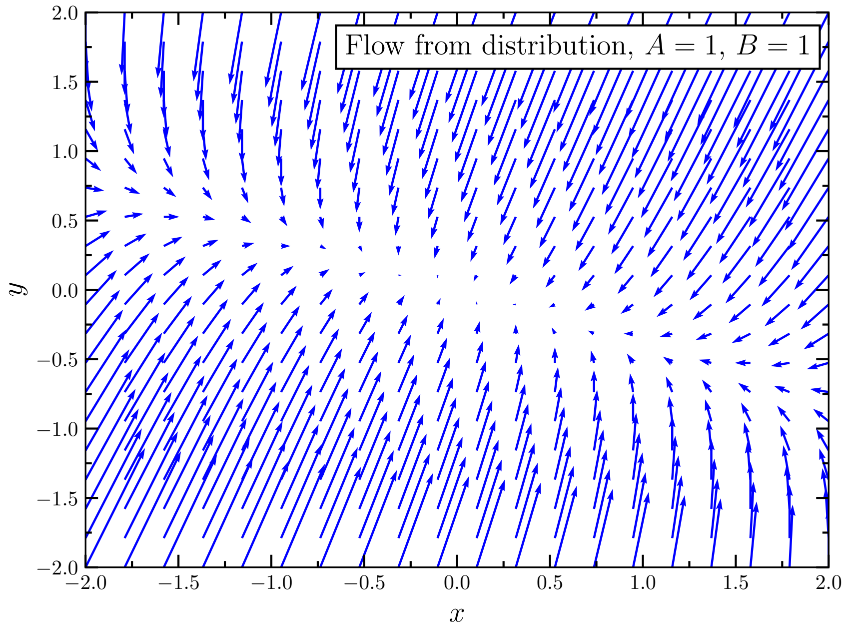
Gaussian model (solvable)

- complex quadratic action $S(x) = \frac{1}{2}\sigma_0 x^2$ $\sigma_0 = A + iB$
- CL equations $\dot{x} = K_x + \eta, \quad K_x = -Ax + By, \quad \dot{y} = K_y, \quad K_y = -Ay - Bx$
- here FPE can be solved $P(x,y) = N \exp\left[-\alpha x^2 - \beta y^2 - 2\gamma xy\right], \qquad N = \frac{1}{\pi}\sqrt{\alpha\beta - \gamma^2}$
- with coefficients $\alpha = A, \beta = A(1 + 2A^2/B^2), \gamma = A^2/B.$
- solution satisfies $\rho(x) = \int dy\, P(x - iy, y)$
- note: score ≠ CL drift $\partial_x \log P(x,y) = -2\alpha x - 2\gamma y, \qquad \partial_y \log P(x,y) = -2\beta y - 2\gamma x$

Flow from CL and from score: Gaussian model

$$A = B = 1$$





CL dynamics:

$$K_x = -Ax + By$$
$$K_y = -Ay - Bx$$

$$\partial_x K_y \neq \partial_y K_x$$

score:

$$\partial_x \log P(x, y) = -2\alpha x - 2\gamma y$$
$$\partial_y \log P(x, y) = -2\beta y - 2\gamma x$$

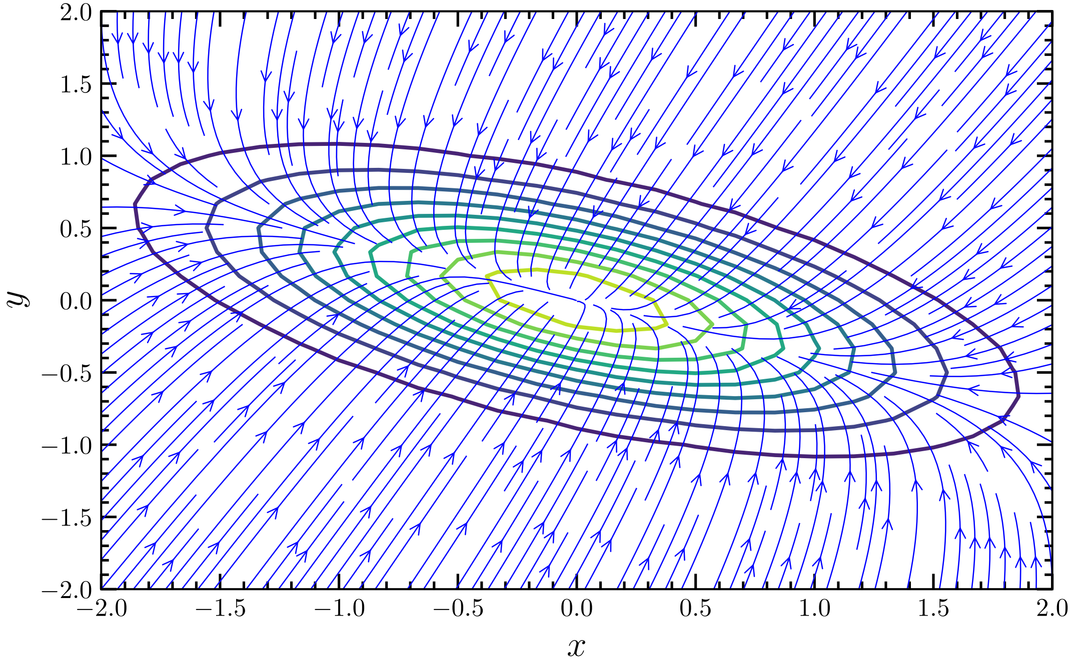
$$\partial_x \partial_y \log P(x, y) = \partial_y \partial_x \log P(x, y)$$

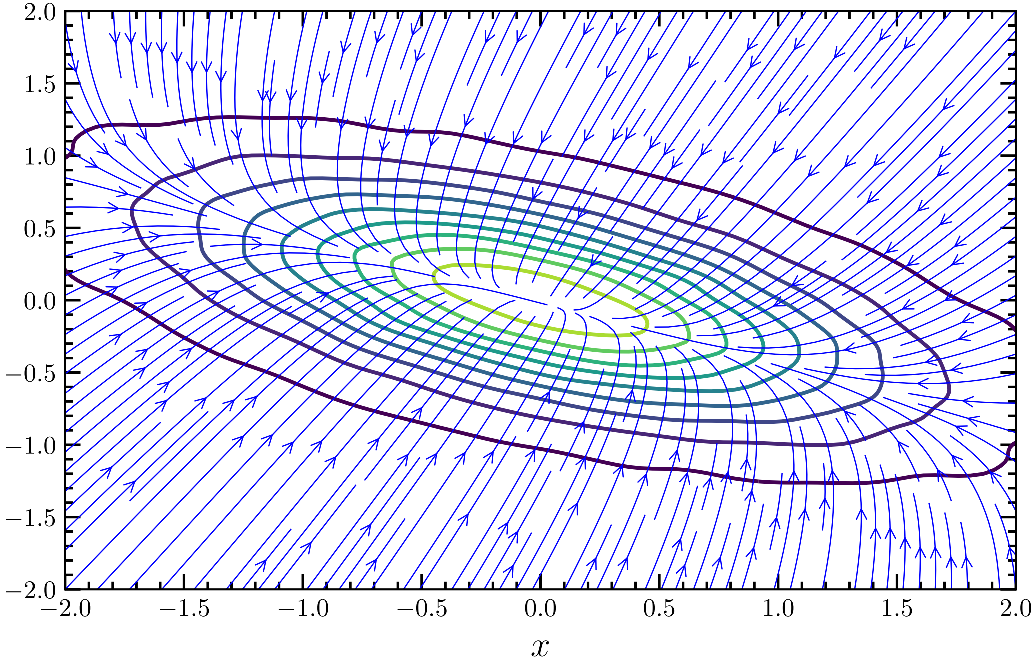
Trained diffusion model: Gaussian case

$A = B = 1$

analytical score

trained model





moments:

n	2		4		6		8	
	re	−im	re	−im	re	−im	re	−im
Exact	0.5	0.5	0	1.5	−3.75	3.75	−26.25	0
CL	0.4986(7)	0.4990(7)	−0.0018(1)	1.494(5)	−3.75(2)	3.75(3)	−26.4(3)	0.20(3)
DM	0.497(1)	0.491(1)	0.021(1)	1.476(7)	−3.65(3)	3.78(4)	−26.3(1)	0.81(68)

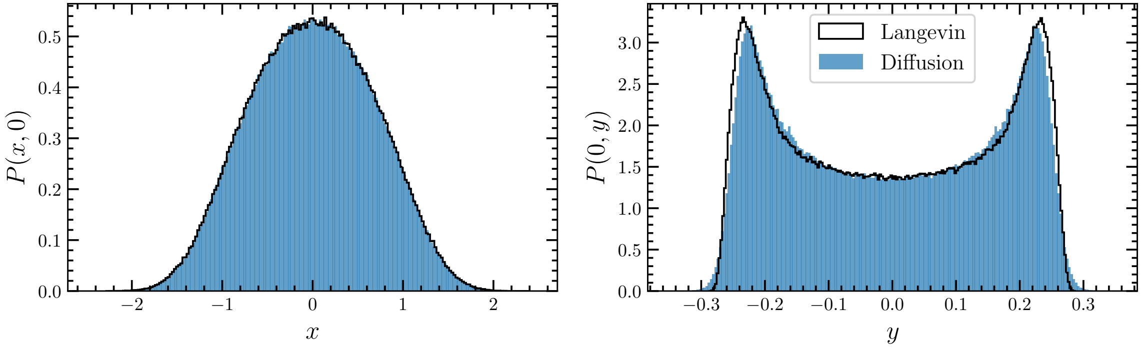
Quartic model

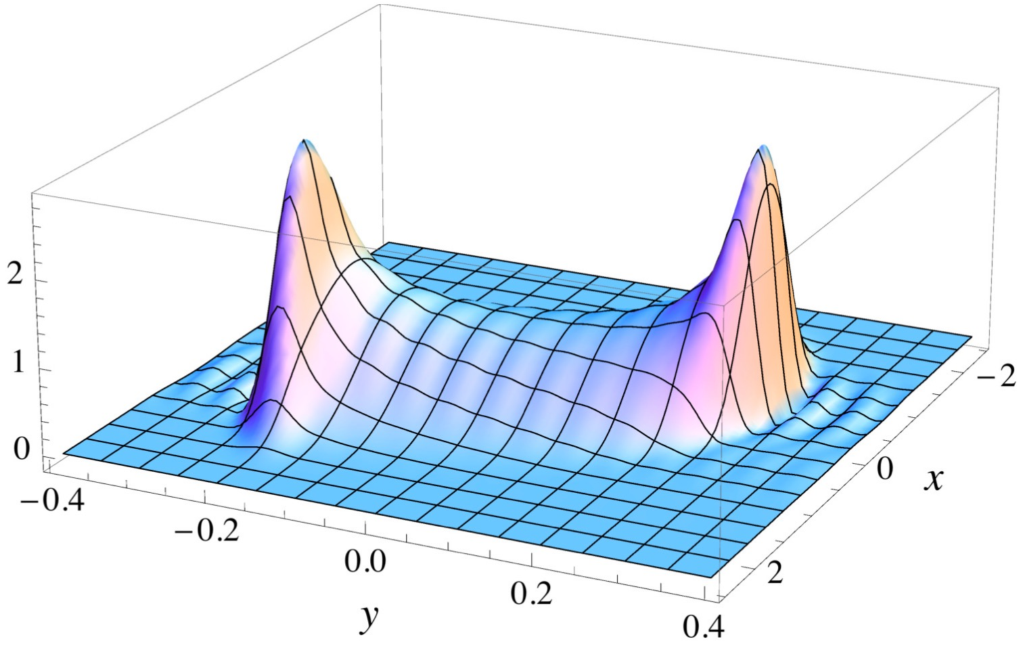
- simple model with quartic coupling $S = \frac{1}{2}\sigma_0 x^2 + \frac{1}{4}\lambda x^4$ $\sigma_0 = A + iB$
- detailed analysis in GA, Giudice, Seiler, *Annals Phys.* **337** (2013) 238 [1306.3075]
- CL converges, provided $3A^2 - B^2 > 0$, dynamics is contained inside a strip, $-y_- < y < y_-$
- this follows from CL drift $$y_-^2 = \frac{A}{2\lambda}\left(1 - \sqrt{1 - \frac{B^2}{3A^2}}\right)$$
- FPE can be solved (approximately) using double expansion in Hermite polynomials
- train diffusion model on CL generated data

Quartic model

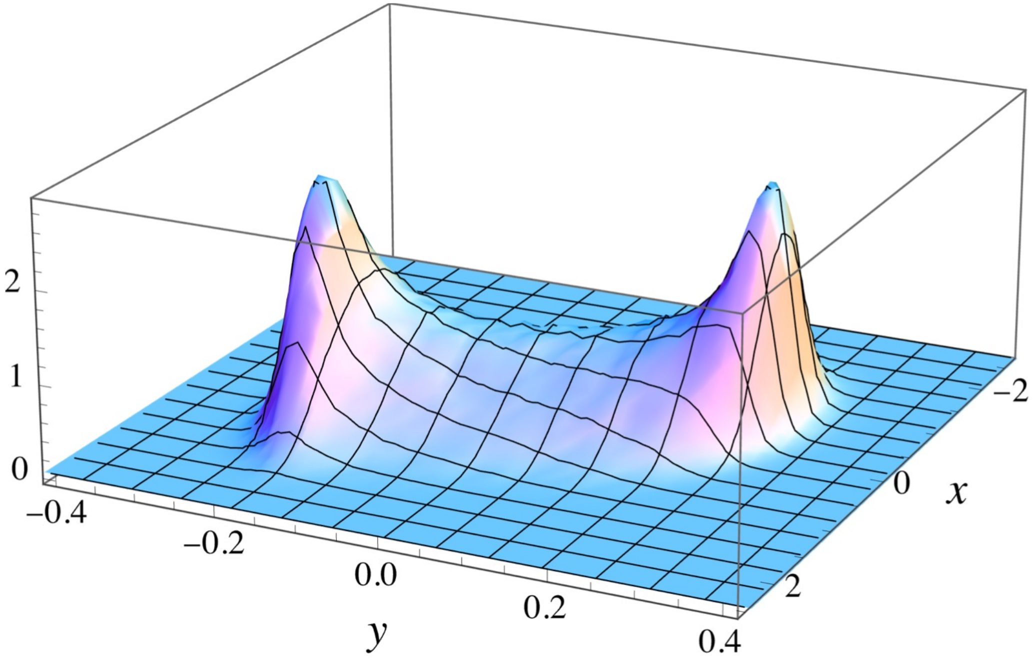
$$A = B = \lambda = 1$$

$$y_- \approx 0.3029$$





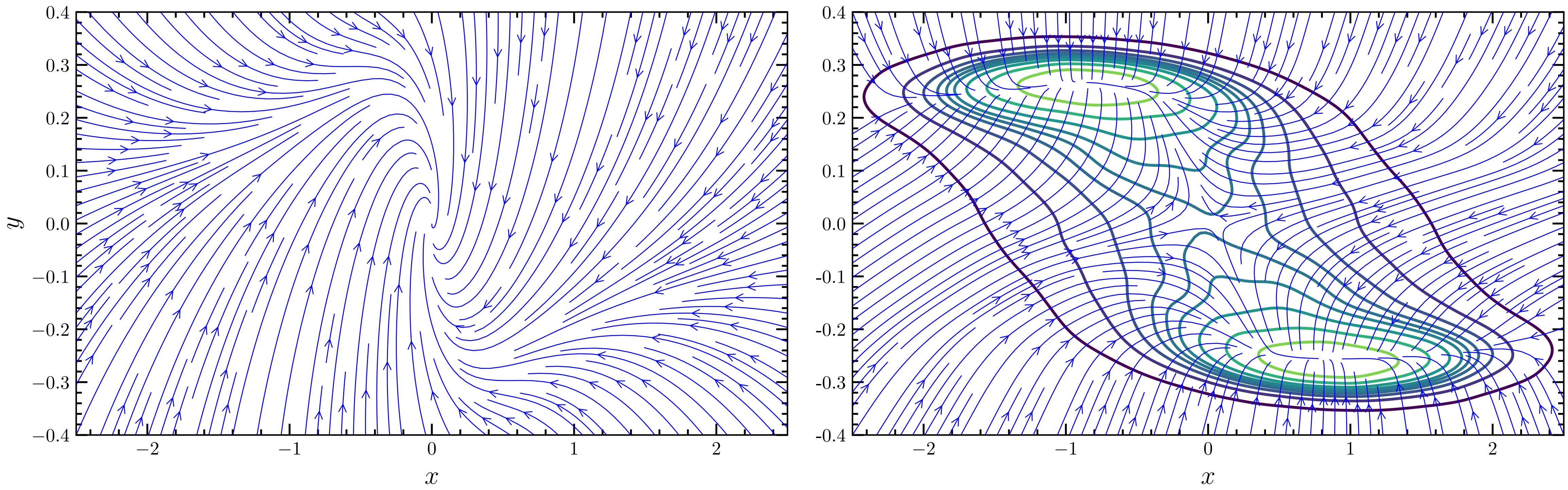
solution of FPE using double expansion in Hermite polynomials



solution obtained by sampling from trained diffusion model

GA, Giudice, Seiler, *Annals Phys.* **337** (2013) 238 [1306.3075]

Trained diffusion model: quartic model



complex Langevin drift

score from trained diffusion model

$$A = B = \lambda = 1$$

$$y_- \approx 0.3029$$

Comparison

cumulants in the quartic model

n	2		4		6		8	
	re	−im	re	−im	re	−im	re	−im
Exact	0.428142	0.148010	−0.060347	−0.100083	−0.00934	0.19222	0.41578	−0.5923
CL	0.4277(5)	0.1478(2)	−0.0597(6)	−0.0991(6)	−0.010(1)	0.188(2)	0.406(4)	−0.57(1)
DM	0.4267(6)	0.1459(2)	−0.0582(6)	−0.0981(5)	−0.008(1)	0.188(2)	0.400(5)	−0.58(1)

expectation values at the end of the backward process

note: diffusion model learns from CL data, not the “exact” value

Trained diffusion model: quartic model

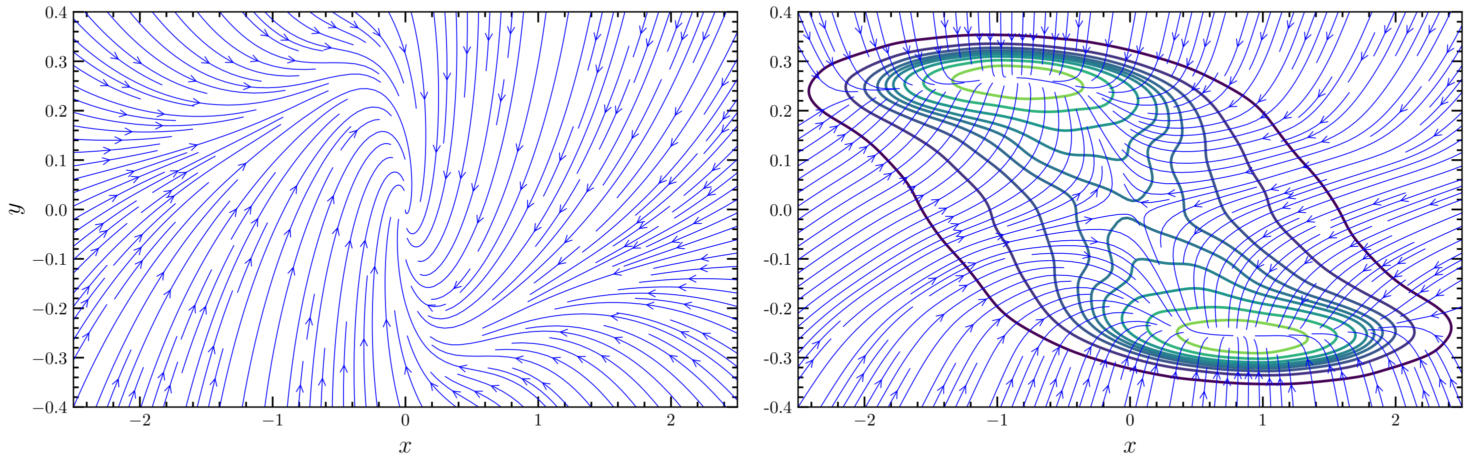
very different processes

complex Langevin:

- non-integrable drift
- noise in real direction
- attractor at origin

diffusion model:

- integrable score
- noise in both directions
- saddle at origin



different Fokker-Planck equations

yet same distributions are created for data generation

have obtained access to $\nabla \log P(x, y)$

Summary and outlook

- diffusion models offer a new approach for ensemble generation to explore in LFT
- learn from data: requires high-quality ensembles
- close relation to stochastic quantisation
- moment- and cumulant-generating functionals

  higher n-point functions important in LFT applications

- apply to complex actions/complex Langevin: DMs learn elusive real-valued distributions
- in progress: apply to theories with fermions

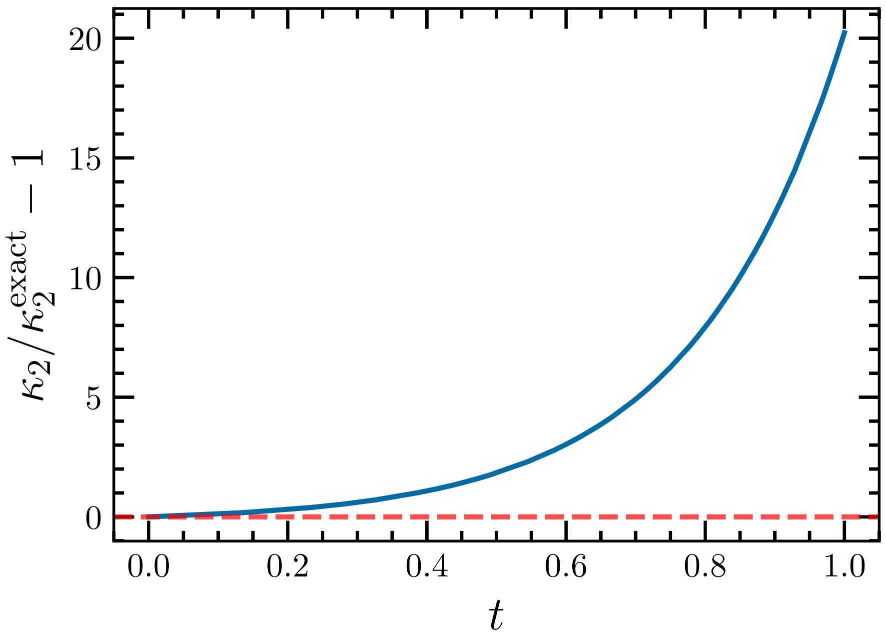
  auxiliary field bosonic models, or DMs learn presence of fermions implicitly

BACKUP SLIDES

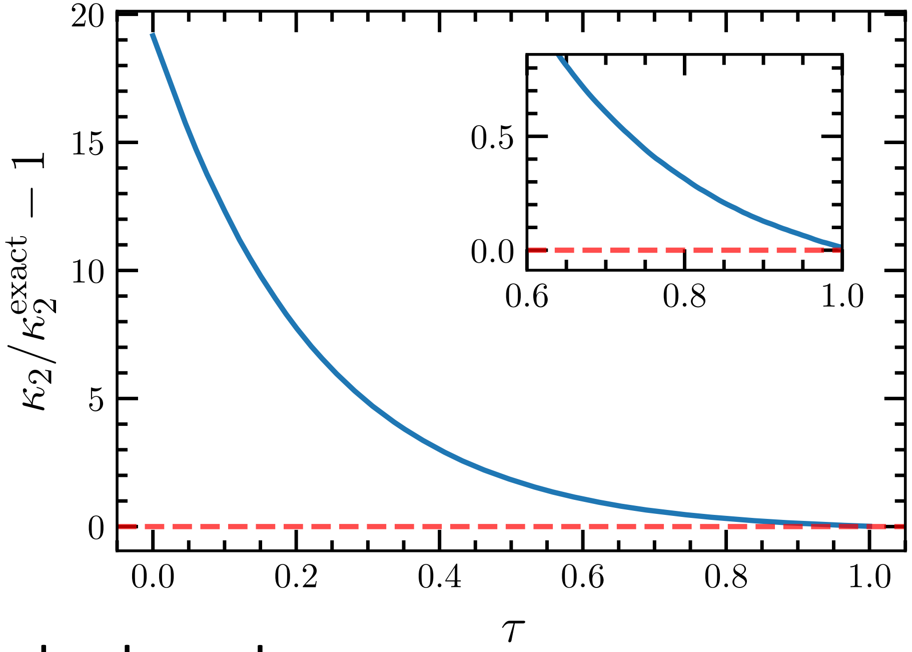
$f(t,s)=1$

2nd cumulant without drift

- variance-expanding scheme $\kappa_2(t)=\kappa_2(0)+\Xi(t)$ $\Xi(t)=\int_0^t ds\, g^2(s)\sim\sigma^{2t/T}$



forward



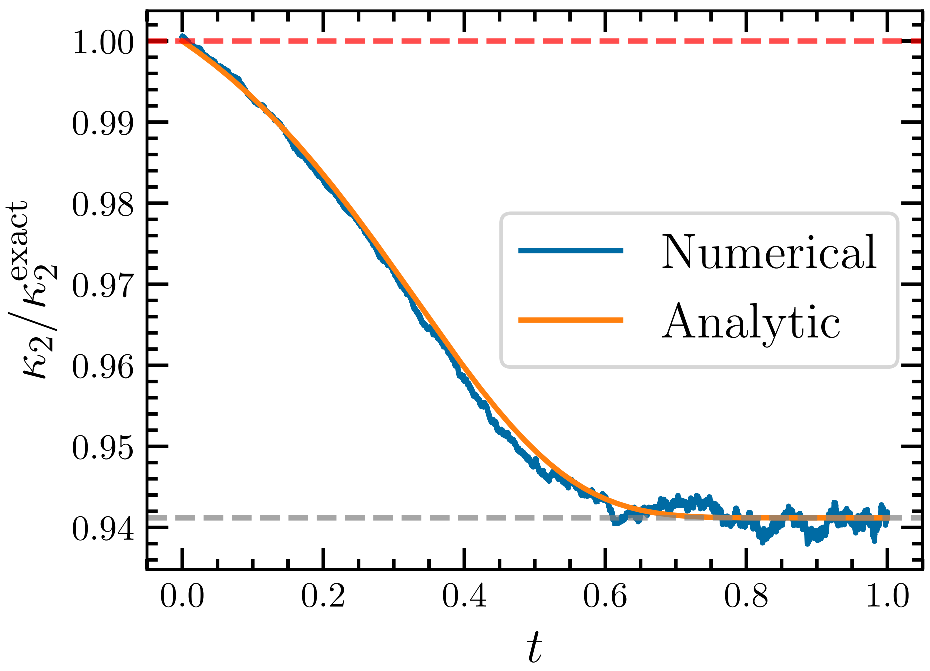
backward

$$f(t,s) = e^{-\frac{1}{2}u(t)+\frac{1}{2}u(s)}$$

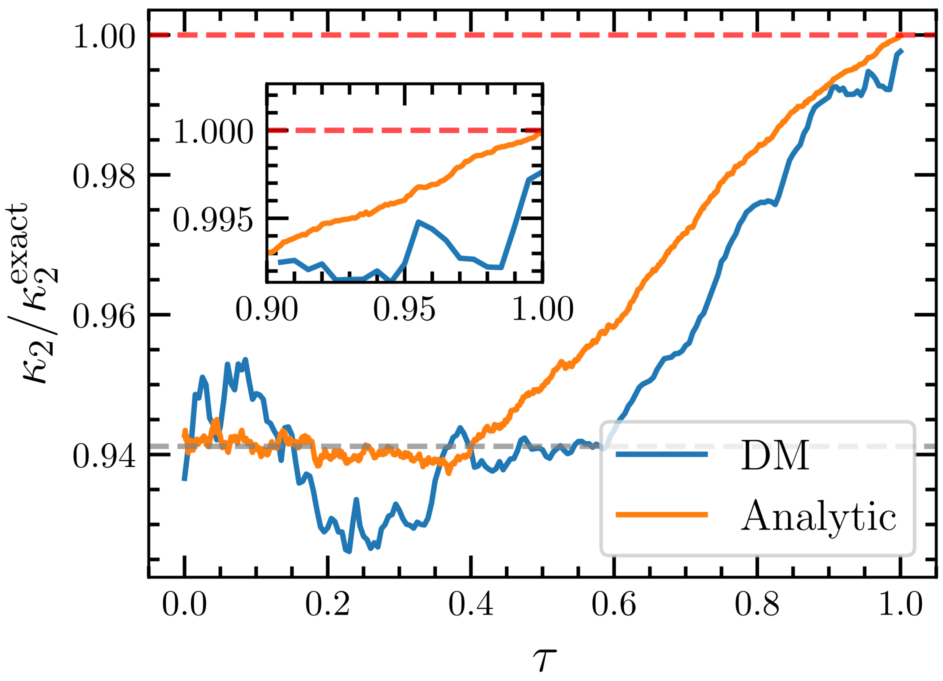
$$u(t) = \int_0^t ds\, g^2(s)$$

2nd cumulant with drift (DDPM)

- variance-preserving scheme $\kappa_2(t) = \mu^2(t) + \sigma^2(t) = \left(\mu_0^2 + \sigma_0^2 - 1\right) f^2(t,0) + 1$



forward



backward

analytic = analytic score