Diffusion models: cumulants and lattice field theory

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

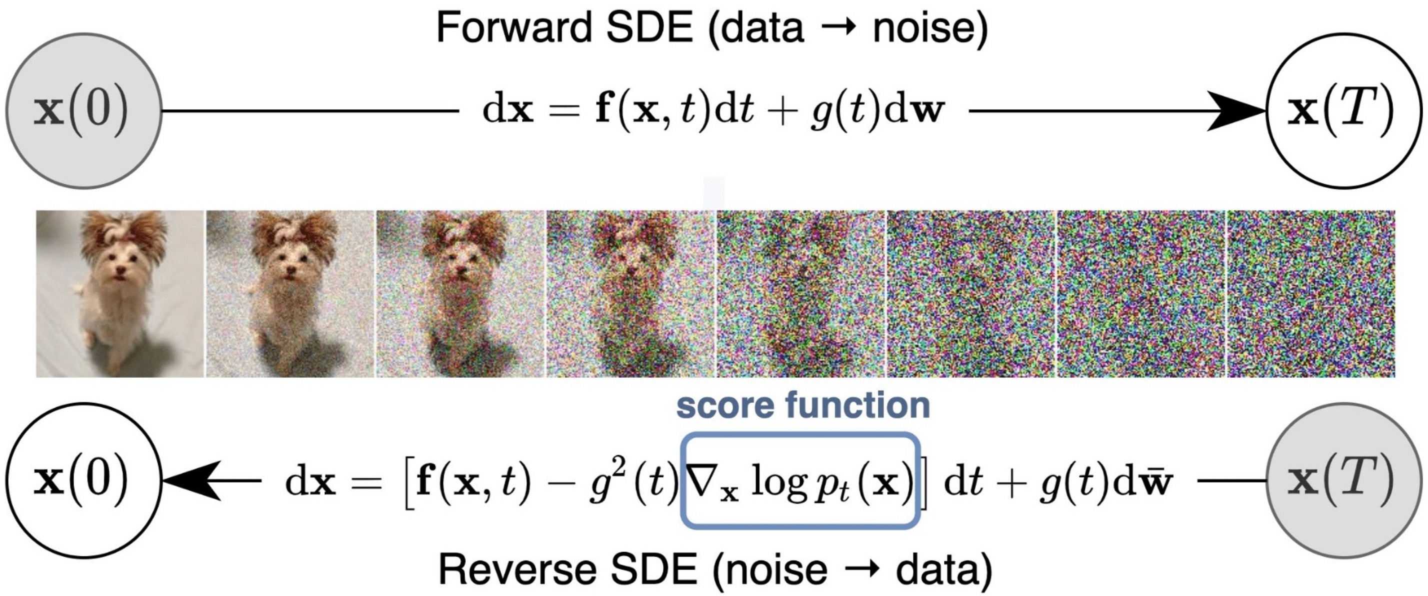
LFT and ML, NTU, Taiwan, Dec 2024

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

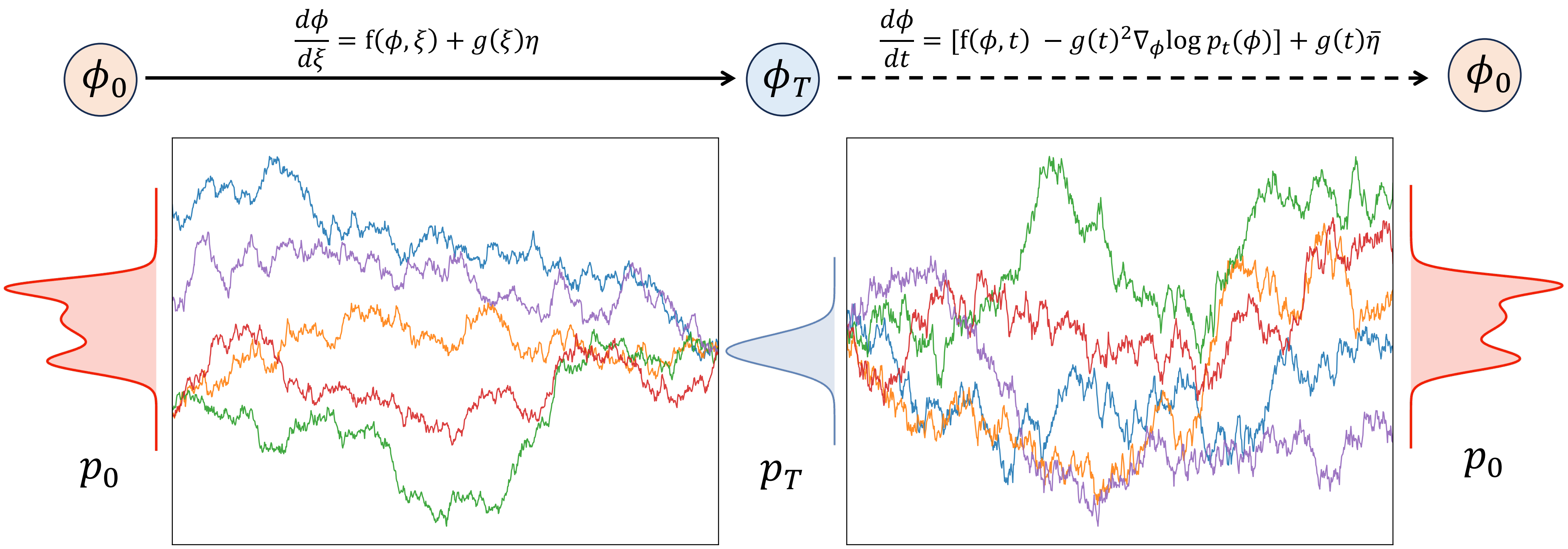
  … and then revert it
- create new images from noise



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

Prior and target distributions

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



forward process

backward process

Outline

- some comments on diffusion models and stochastic quantisation
- first application in lattice scalar field theory in two dimensions
- correlations: higher n-point functions and interactions in field theory
- detailed analysis of forward and backward process, cumulants, generating functionals
- application to 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)

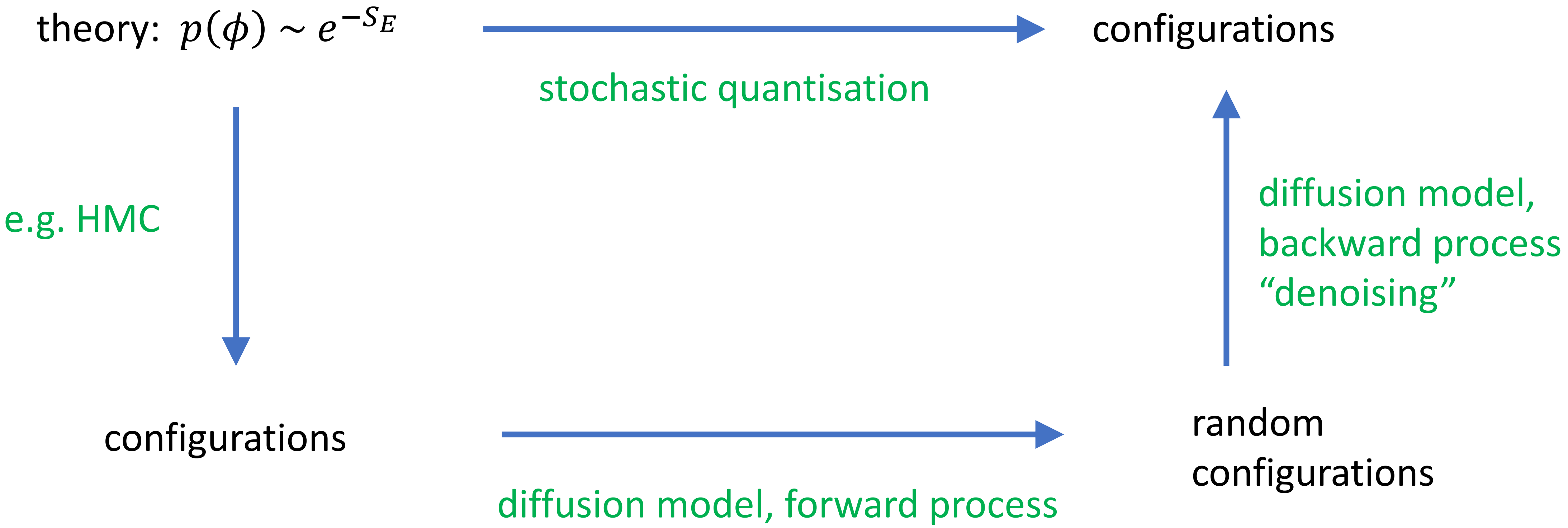
thermalisation followed by long-term evolution in equilibrium

✓ 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



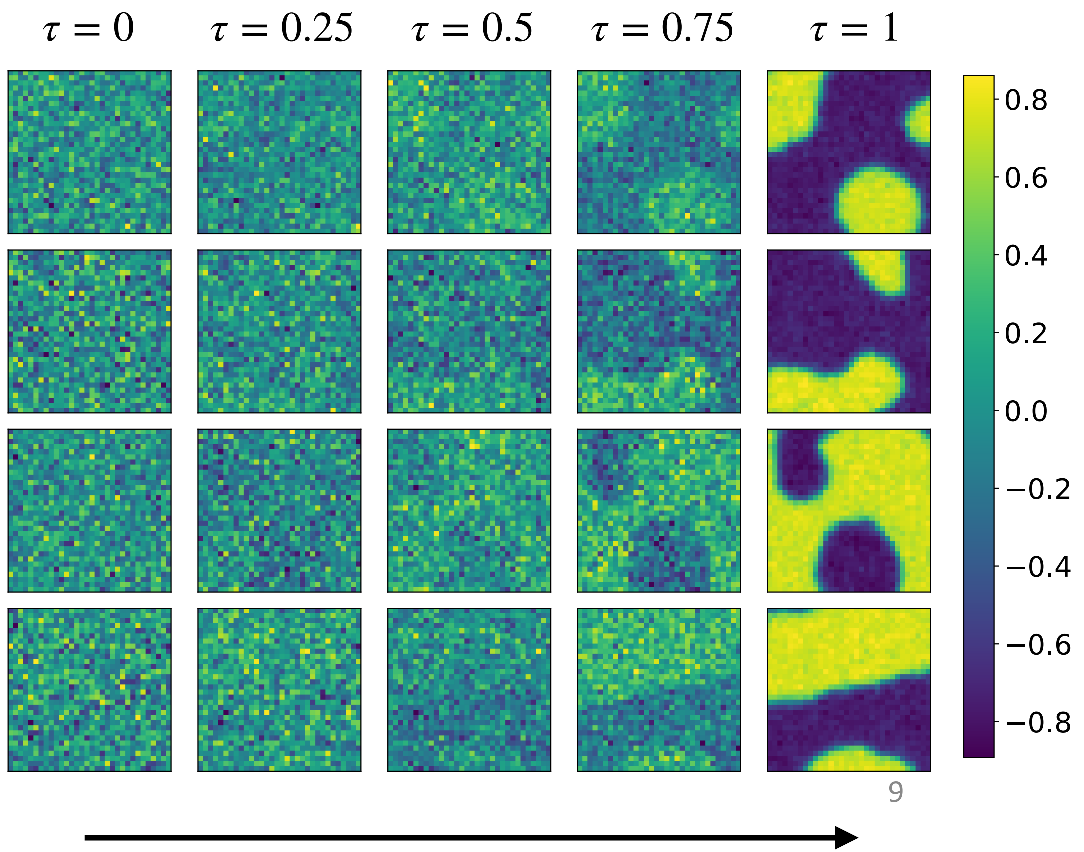
Diffusion model for 2d ϕ^4 scalar theory

- 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

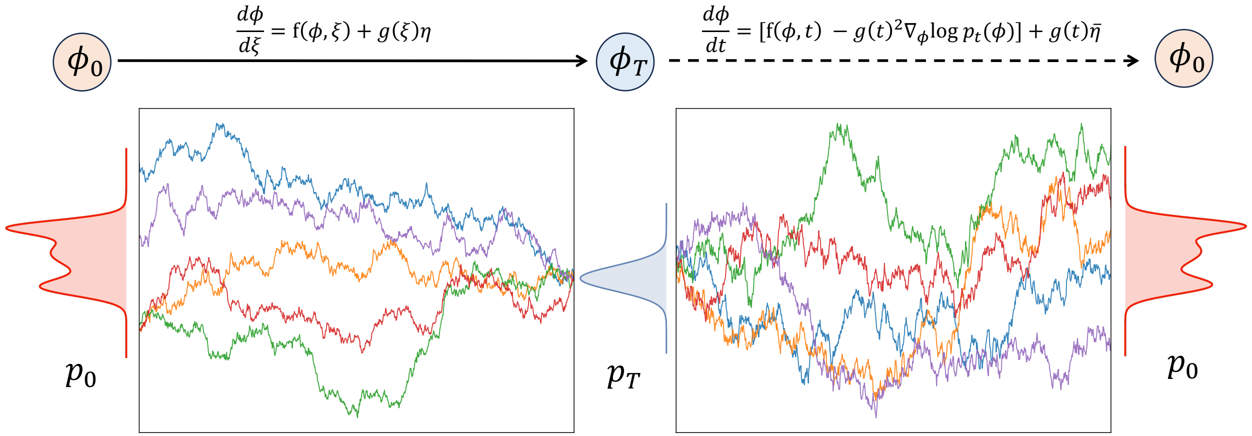
generating configurations:

- 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?
- often stated data at the end of forward process is decorrelated (normal distribution)
- higher n-point functions contain interactions in field theory
- essential for applications in field theory, correlations = interactions
- focus on moments and cumulants
- various schemes/implementations available: (dis)advantages?

discuss forward and backward process in more detail

Diffusion models

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

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

Diffusion models: forward process

- forward process: $\dot{x}(t) = K(x(t), t) + g(t)\eta(t)$ $\quad 0 \le t \le T$
- linear (or zero) drift: $K(x(t), t) = -\frac{1}{2}k(t)x(t)$ $\quad$ noise profile $g(t) = \sigma^{t/T}$
- initial data from target ensemble $x_0 \sim P_0(x_0)$
- solution: $$x(t) = x_0 f(t, 0) + \int_0^t ds\, f(t, s) g(s) \eta(s)$$
- with $$f(t, s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$$

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

Diffusion models: forward process

- solution: $$x(t) = x_0 f(t,0) + \int_0^t ds\, f(t,s) g(s) \eta(s)$$
- moments $\mu_n(t) = \mathbb{E}[x^n(t)]$ and cumulants or connected n-point functions $\kappa_n(t)$
- second moment/cumulant: (assume: first moment vanishes: $x_0 \to x_0 - \mathbb{E}_{P_0}[x_0]$)

$$\kappa_2(t) = \mu_2(t) = \mu_2(0) f^2(t,0) + \Xi(t)$$

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

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

$$\mu_n(t) = \mathbb{E}[x^n(t)]$$

Diffusion models: forward process

- solution: $$x(t) = x_0 f(t,0) + \int_0^t ds\, f(t,s) g(s) \eta(s)$$
- higher-order moment and cumulants:

$$\kappa_2(t) = \mu_2(t) = \mu_2(0) f^2(t,0) + \Xi(t)$$

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

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

$$\kappa_4(t) = \mu_4(t) - 3\mu_2^2(t)$$

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

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

variance-expanding scheme: no drift

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

higher cumulants conserved!

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

Diffusion models: forward process

- higher-order cumulants: $\kappa_{n>2}(t) = \kappa_n(0) f^n(t,0)$

- in variance-expanding scheme ($f(t,0) = 1$, no drift): distribution at end of forward process as correlated as target distribution

- proof to all orders: generating functionals $Z[J] = \mathbb{E}[e^{J(t)x(t)}]$ $\qquad W[J] = \log Z[J]$

- average over both noise and target distribution

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

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

Diffusion models: generating functionals

- noise average: $$Z_\eta[J] = e^{J(t)x_0 f(t,0) + \frac{1}{2}J^2(t)\Xi(t)}$$
- total average: $$Z[J] = \mathbb{E}[e^{J(t)x(t)}] = e^{\frac{1}{2}J^2(t)\Xi(t)} \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$
- cumulants: $$W[J] = \log Z[J] = \frac{1}{2}J^2(t)\Xi(t) + \log \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$
- 2nd cumulant: $$\kappa_2(t) = \frac{d^2 W[J]}{dJ(t)^2}\Big|_{J=0} = \Xi(t) + \mathbb{E}_{P_0}[x_0^2] f^2(t,0) \quad \checkmark$$
- higher-order cumulants: $$\kappa_{n>2}(t) = \frac{d^n W[J]}{dJ(t)^n}\Big|_{J=0} = \frac{d^n}{dJ(t)^n} \log \mathbb{E}_{P_0}[e^{J(t)x_0 f(t,0)}]\Big|_{J=0} = \kappa_n(0) f^n(t,0) \quad \checkmark$$

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

Diffusion models: generating functionals

- exact expression for cumulant-generating function
  (for any linear or vanishing drift and noise strength)

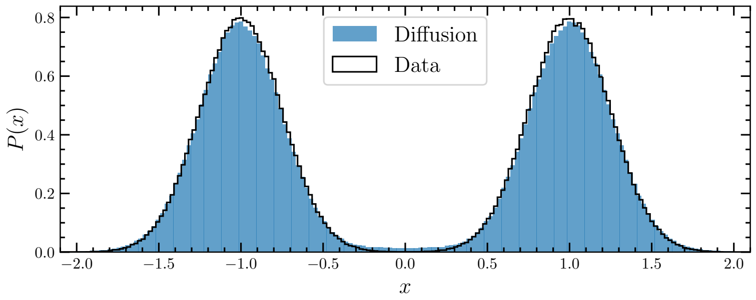
$$W[J] = \log Z[J] = \frac{1}{2} J^2(t)\Xi(t) + \log \int dx_0\, P_0(x_0) e^{J(t) x_0 f(t,0)}$$

- particularly interested in higher-order cumulants

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

- apply/test in simple model and lattice field theory

Toy model: two-peak distribution



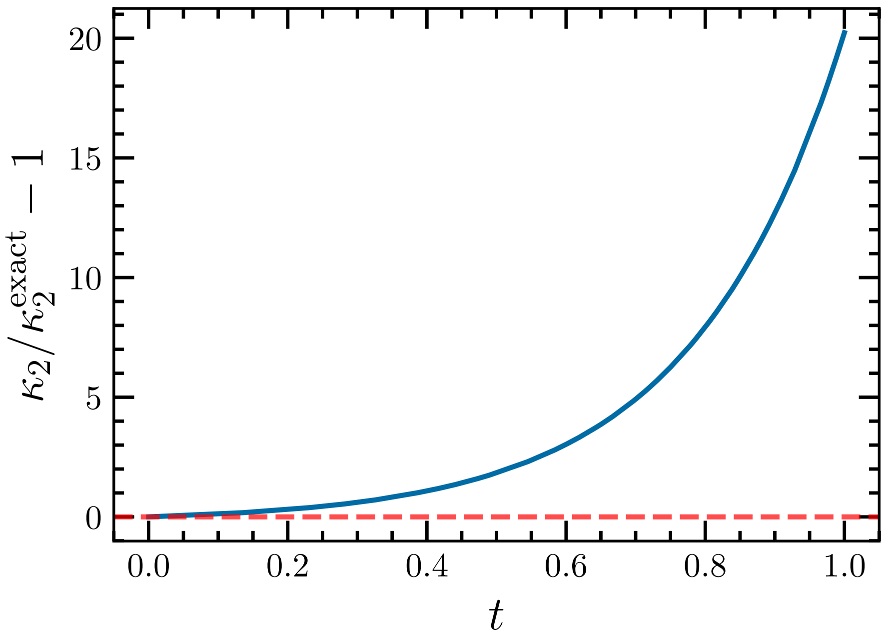
- sum of two Gaussians: $$P_0(x) = \frac{1}{2}\left[\mathcal{N}(x;\mu_0,\sigma_0^2) + \mathcal{N}(x;-\mu_0,\sigma_0^2)\right]$$
- moment-generating function: $$Z[j] = \mathbb{E}\left[e^{jx}\right] = e^{\frac{1}{2}\sigma_0^2 j^2}\cosh(\mu_0 j)$$
- cumulant-generating function: $$W[j] = \frac{1}{2}\sigma_0^2 j^2 + \log\cosh(\mu_0 j)$$
- only second cumulant depends on σ_0^2 :

$$\kappa_2 = \mu_0^2 + \sigma_0^2, \qquad \kappa_4 = -2\mu_0^4, \qquad \kappa_6 = 16\mu_0^6, \qquad \kappa_8 = -272\mu_0^8$$ etc

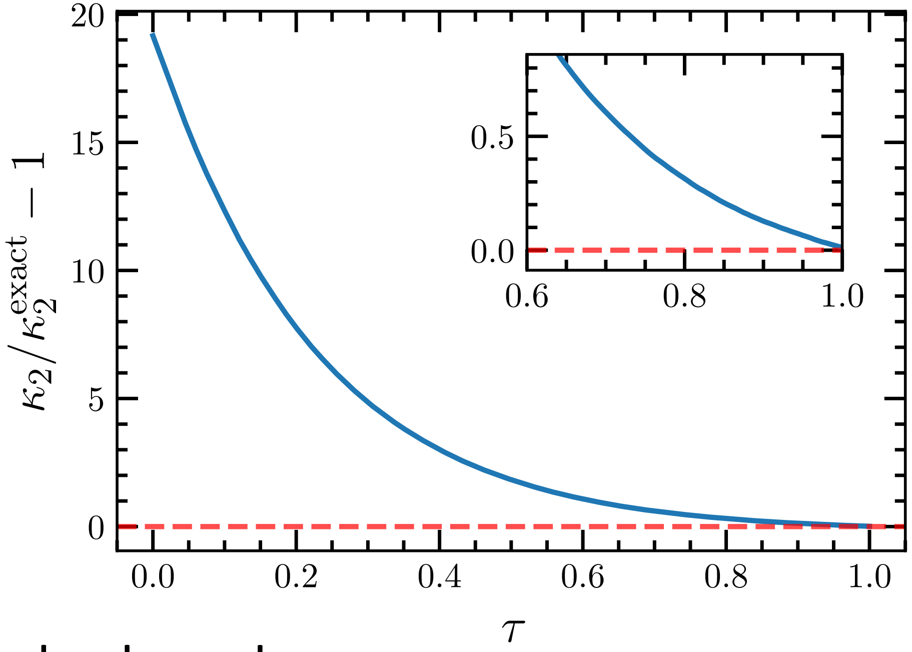
$f(t, s) = 1$

2nd cumulant without drift

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



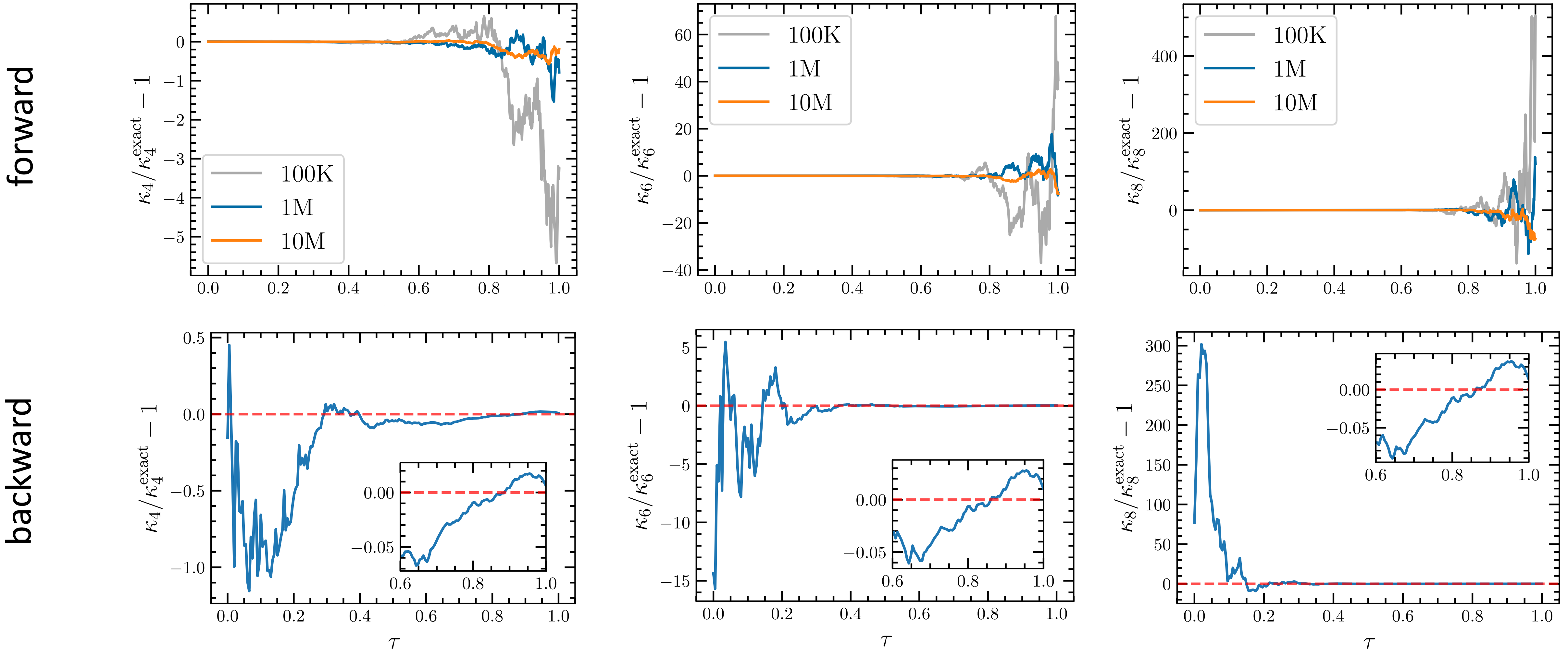
forward



backward

$f(t,s)=1$

4^{th}, 6^{th}, 8^{th} cumulant without drift $\kappa_{n>2}(t)=\kappa_n(0)$



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

- 2^{nd} cumulant increases as expected: variance expanding
- higher-order cumulants are conserved, up to numerical cancellations:
  cumulants require cancellations between moments which increase in time
- initial conditions for backward process taken from normal distribution:
  higher-order cumulants initially vanish, up to numerical cancellations
- score has higher-order cumulants encoded: cumulants are reconstructed

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

- score has higher-order cumulants encoded: cumulants are reconstructed
- how do we know this (besides numerical evidence)?
- time-dependent distribution and score can be given analytically:

$$P(x,t) = \frac{1}{2}\left[\mathcal{N}(x;\mu_0,\sigma^2(t)) + \mathcal{N}(x;-\mu_0,\sigma^2(t))\right]$$

- with $\sigma^2(t) = \sigma_0^2 + \Xi(t)$ $\Xi(t) = \int_0^t ds\, g^2(s) \sim \sigma^{2t/T}$

- score: $-\partial_x \log P(x,t) = \frac{x}{\sigma^2(t)} - \frac{\mu_0}{\sigma^2(t)} \tanh\left(\frac{\mu_0 x}{\sigma^2(t)}\right)$

- encodes all information about higher-order cumulants (solve process with this score ✓)

DDPM: with drift

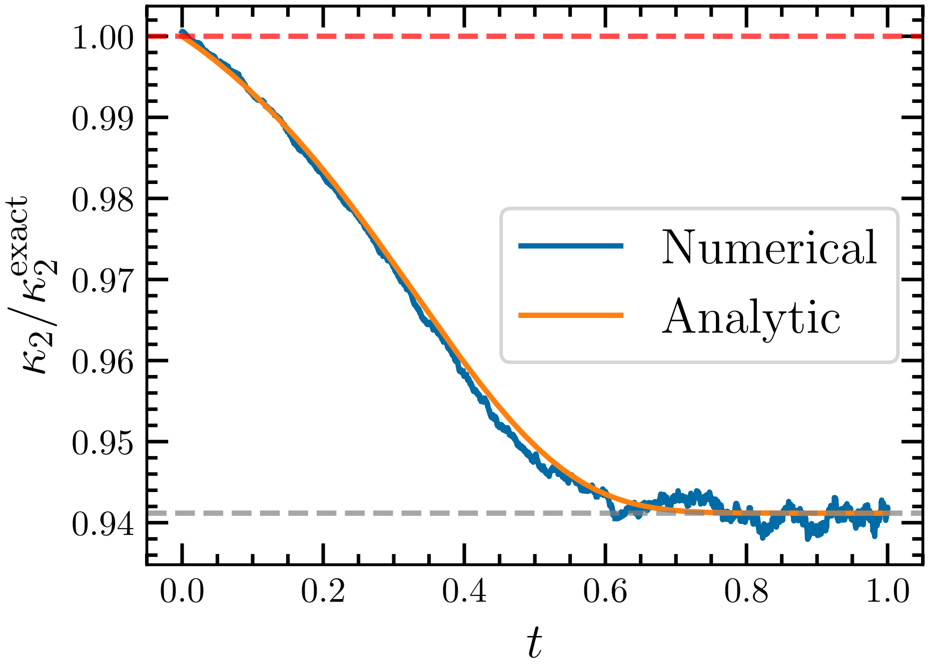
- include a linear drift $K(x(t),t) = -\frac{1}{2}k(t)x(t)$
- choice of coefficient $k(t) = g^2(t)$
- simple FPE $\partial_t P(x,t) = \frac{1}{2}g^2(t)\partial_x\left(\partial_x + x\right)P(x,t)$
- redefine time $u(t) = \int_0^t ds\, g^2(s)$
- simplest FPE $\partial_u P(x,u) = \frac{1}{2}\partial_x\left(\partial_x + x\right)P(x,u)$

2^{nd} cumulant with drift (DDPM)

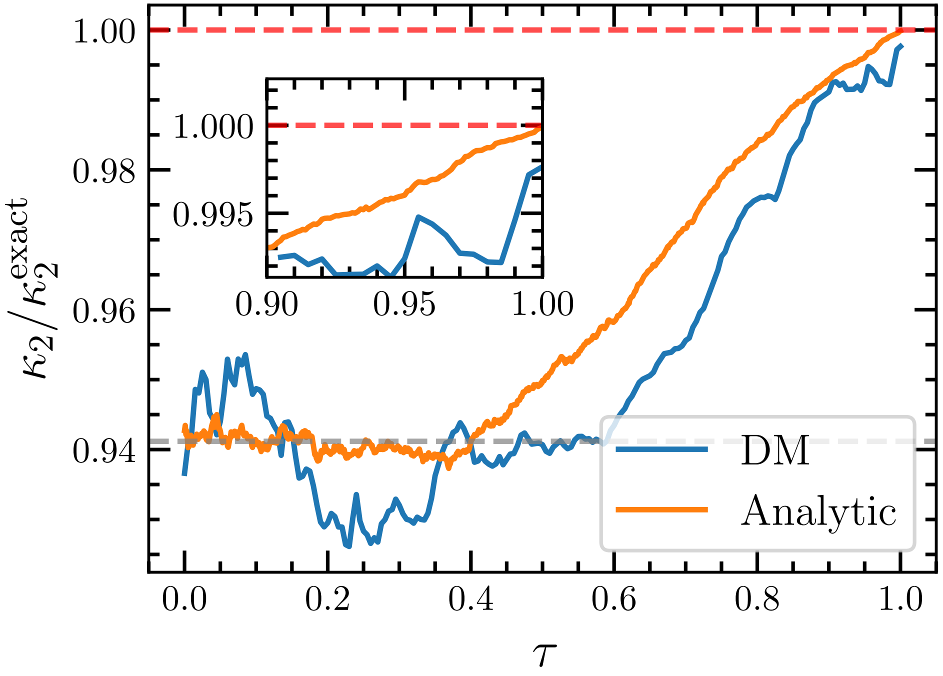
$$f(t,s) = e^{-\frac{1}{2}u(t)+\frac{1}{2}u(s)}$$

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

- 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

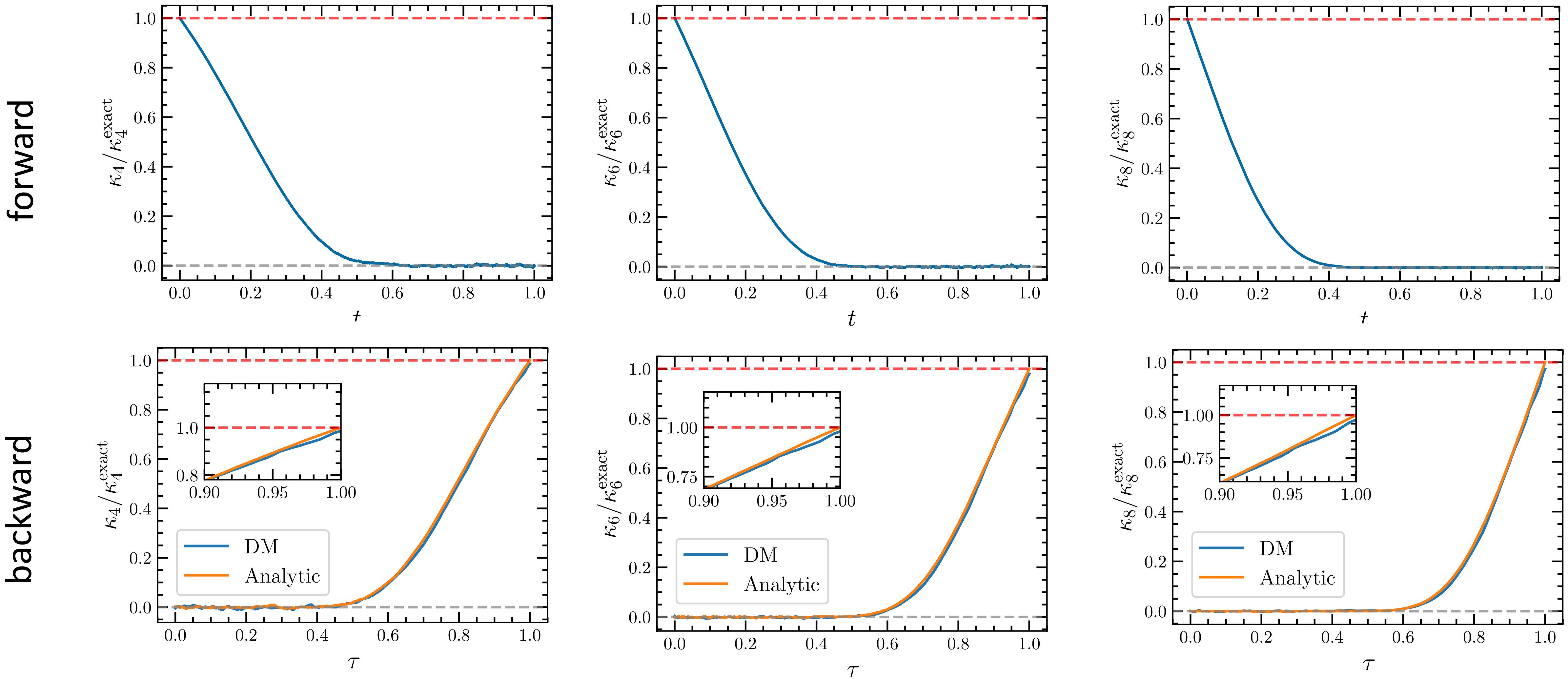


backward

analytic = analytic score

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

4^{th}, 6^{th}, 8^{th} cumulant with drift (DDPM)



analytic = analytic score

2^{nd}, 4^{th}, 6^{th}, 8^{th} cumulant with drift (DDPM)

- 2^{nd} cumulant goes to 1: variance preserving (but noisy!)
- higher-order cumulants go to zero → distribution becomes normal indeed
  no numerical cancellations required
- initial conditions for backward process taken from normal distribution
- cumulants interpolate smoothly
- score has higher-order cumulants encoded: cumulants are reconstructed

2^{nd}, 4^{th}, 6^{th}, 8^{th} cumulant with drift (DDPM)

- time-dependent distribution and score can be given analytically:

$$P(x,t) = \frac{1}{2}\left[\mathcal{N}(x;\mu(t),\sigma^2(t)) + \mathcal{N}(x;-\mu(t),\sigma^2(t))\right]$$

- with $\sigma^2(t) = \sigma_0^2 f^2(t,0) + \Xi(t)$

$$\mu(t) = \mu_0 f(t,0)$$

- encodes all information about higher-order cumulants (solve process with this score ✓)

$$\Xi(t) = \int_0^t ds\, f^2(t,s) g^2(s) = 1 - f^2(t,0)$$

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

$$u(t) = \int_0^t ds\, g^2(s) = \frac{T}{\log\sigma^2}\left[\sigma^{2t/T} - 1\right]$$

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

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: $$\mu_n(x,t) = \mathbb{E}[\phi^n(x,t)]$$

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:

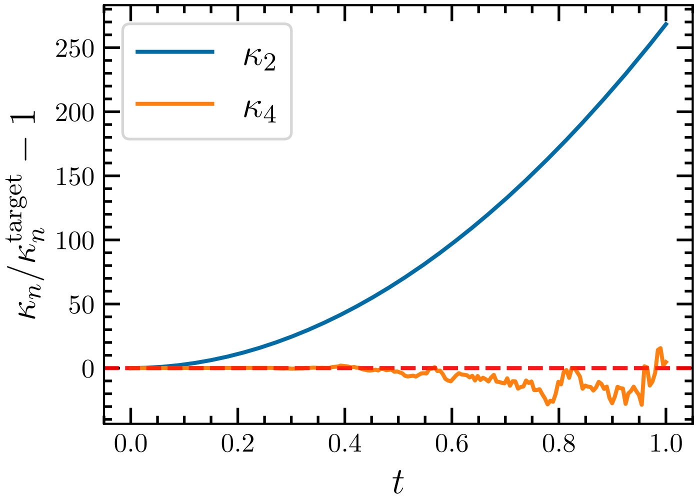
$$\kappa_{n>2}(t) = \frac{\delta^n W[J]}{\delta J(x,t)^n}\Big|_{J=0} = \frac{\delta^n}{\delta J(x,t)^n} \log \mathbb{E}_{P_0}[e^{J(x,t)\phi_0(x)f(t,0)}]\Big|_{J=0}$$

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

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

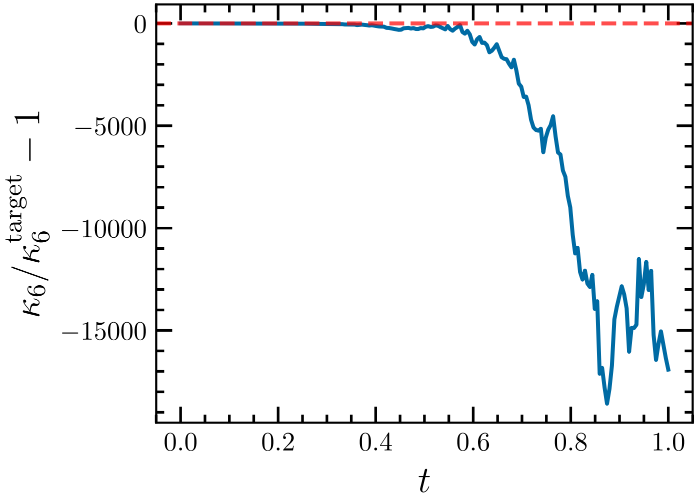
2^{nd}, 4^{th}, 6^{th} cumulant without drift

forward

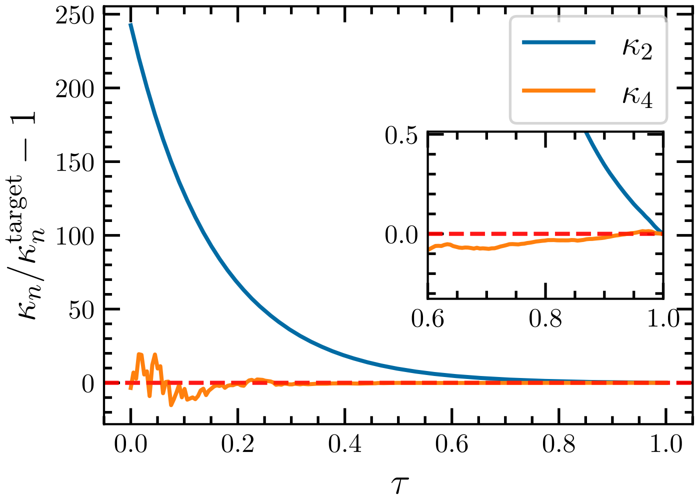


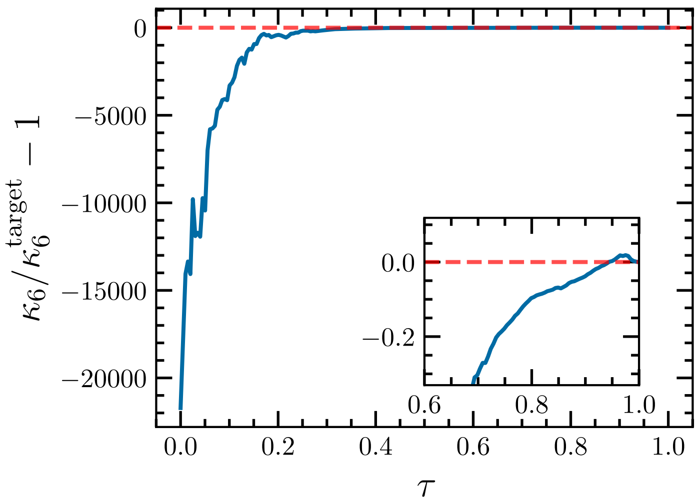
κ_2, κ_4

κ_6



backward





Comparison

	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)

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

expectation values at the end of the backward process

excellent agreement

Summary of part I

- dynamics in diffusion models described in terms of generating functionals
- clarifies evolution of cumulants
- quite general description for linear or vanishing drift
- interpolate between variance-expanding and variance-preserving scheme (not shown)

Complex actions 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)
- recent applications in QCD (Sexty et al, 2023, 2024)
- introductory lectures (GA, 1512.05145 [hep-lat])

(Complex) Langevin dynamics

- observables: $\langle O(x)\rangle = \int dx\, \rho(x) O(x), \qquad \rho(x) = \frac{1}{Z}\exp[-S(x)], \qquad Z = \int dx\, \rho(x)$
- Langevin equation and drift: $\dot{x}(t) = K[x(t)] + \eta(t), \qquad K(x) = \frac{d}{dx}\log\rho(x) = -\frac{dS(x)}{dx}$
- 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: $x \longrightarrow z = x + iy$

Complex Langevin dynamics

- complexify degrees of freedom: $x \longrightarrow z = x + iy$
- Langevin equation and drift: $\dot{z}(t) = K[z(t)] + \eta(t),$ $\quad K(z) = \frac{d}{dz} \log \rho(z) = -\frac{dS(z)}{dz}$
- 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: $\partial_t P(x,y;t) = \left[\partial_x \left(N_x\partial_x - K_x\right) + \partial_y \left(N_y\partial_y - K_y\right)\right] P(x,y;t)$

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

- 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: $$\partial_y K_x \neq \partial_x K_y.$$
- 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

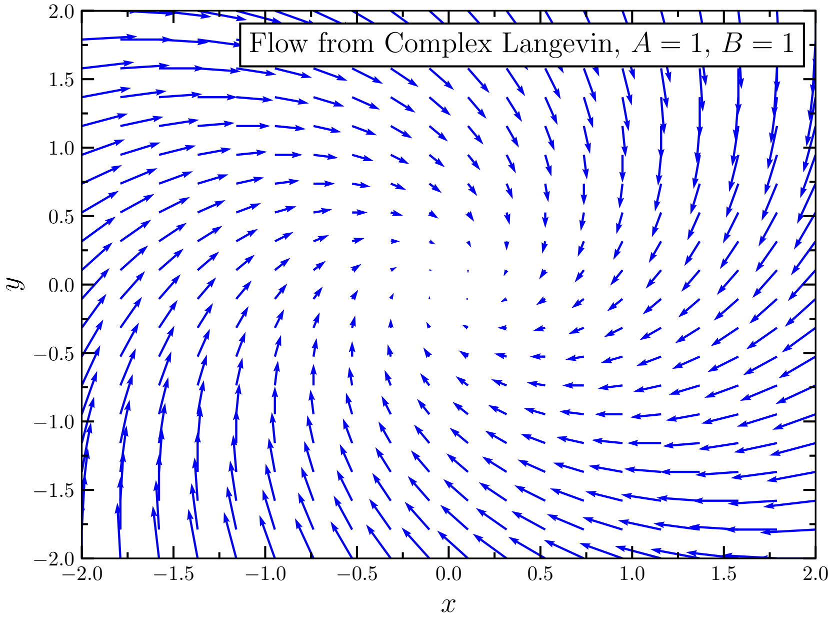
- aside: exact distribution is not known, corrective accept/reject step not possible

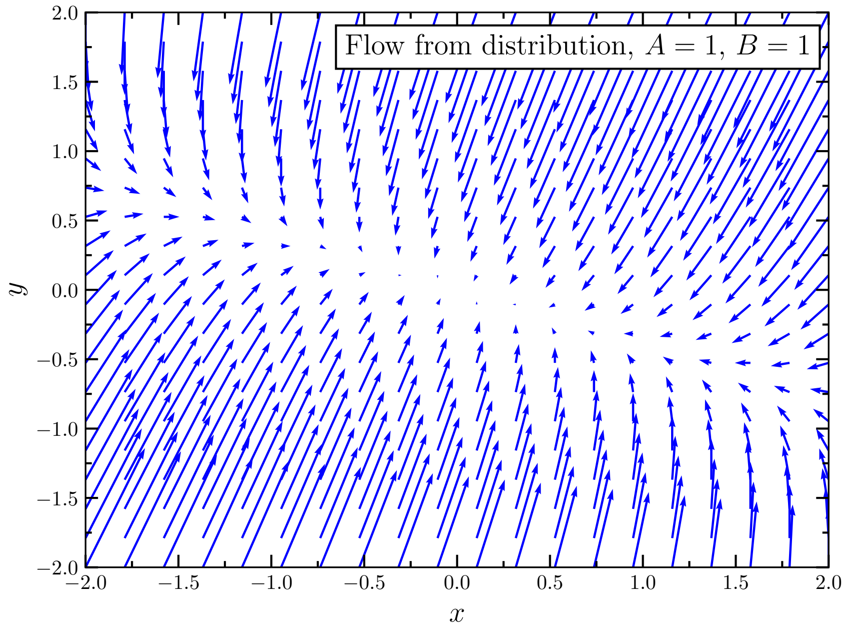
Gaussian model (solvable)

- complex quadratic action: $S(x) = \frac{1}{2}\sigma_0 x^2, \qquad \sigma_0 = A + iB.$
- CL equations: $\dot{x} = K_x + \eta, \qquad K_x = -Ax + By, \qquad \dot{y} = K_y, \qquad K_y = -Ay - Bx$
- 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 $\qquad \partial_x \log P(x,y) = -2\alpha x - 2\gamma y, \qquad \partial_y \log P(x,y) = -2\beta y - 2\gamma x.$

Flow from CL and from score: Gaussian model

$A = B = 1$





CL dynamics:

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

score:

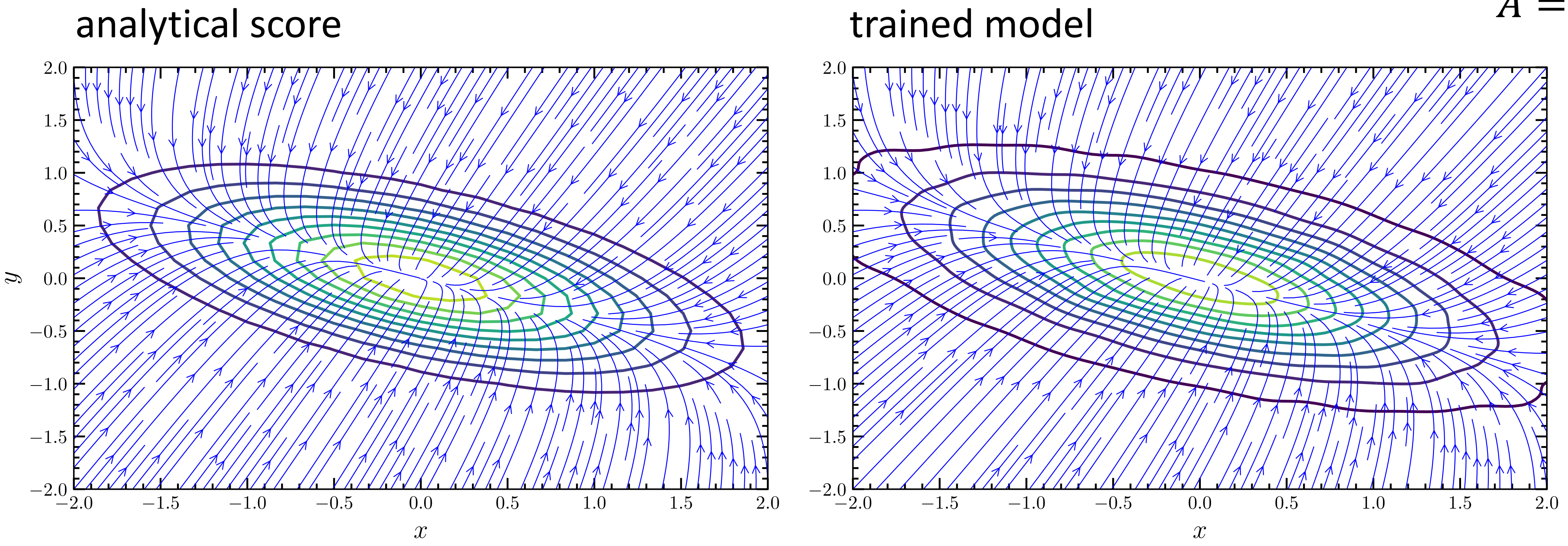
$$\partial_x \log P(x,y) = -2\alpha\, x - 2\gamma\, y$$
$$\partial_y \log P(x,y) = -2\beta\, x - 2\gamma\, x$$

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

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

Trained diffusion model: Gaussian case

$A = B = 1$



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)

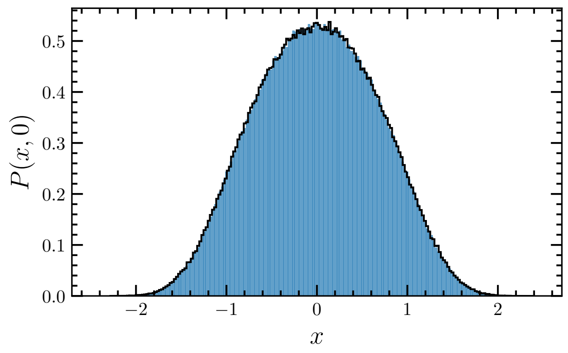
Quartic model

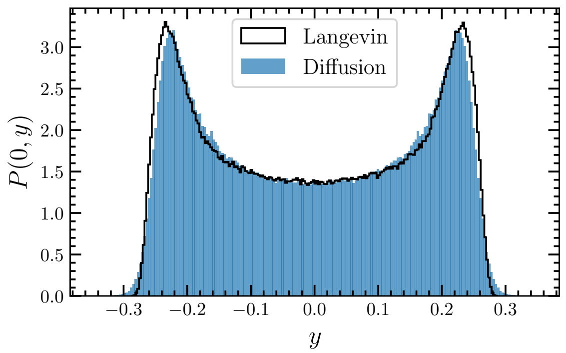
- simple model with quartic coupling $$S = \frac{1}{2}\sigma_0 x^2 + \frac{1}{4}\lambda x^4, \qquad \sigma_0 = A + iB.$$
- detailed analysis in GA, Giudice, Seiler, *Annals Phys.* **337** (2013) 238 [1306.3075]
- CL converges, provided $3A^2 - B^2 > 0$, dynamics is contained inside a strip, $-y_- < y < y_-$
- 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

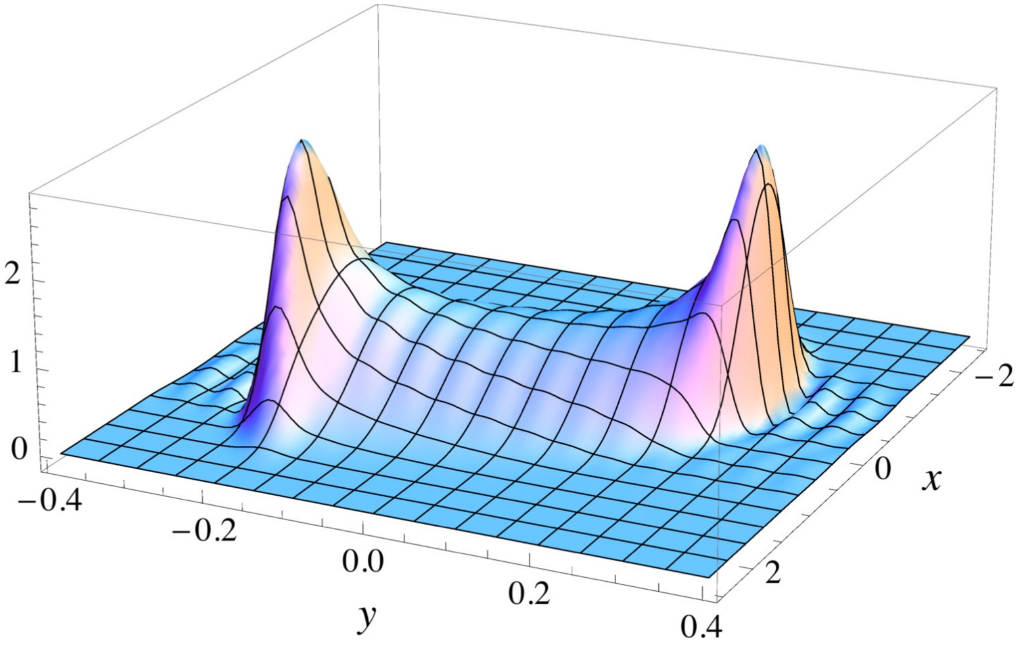
Quartic model

$A = B = \lambda = 1$

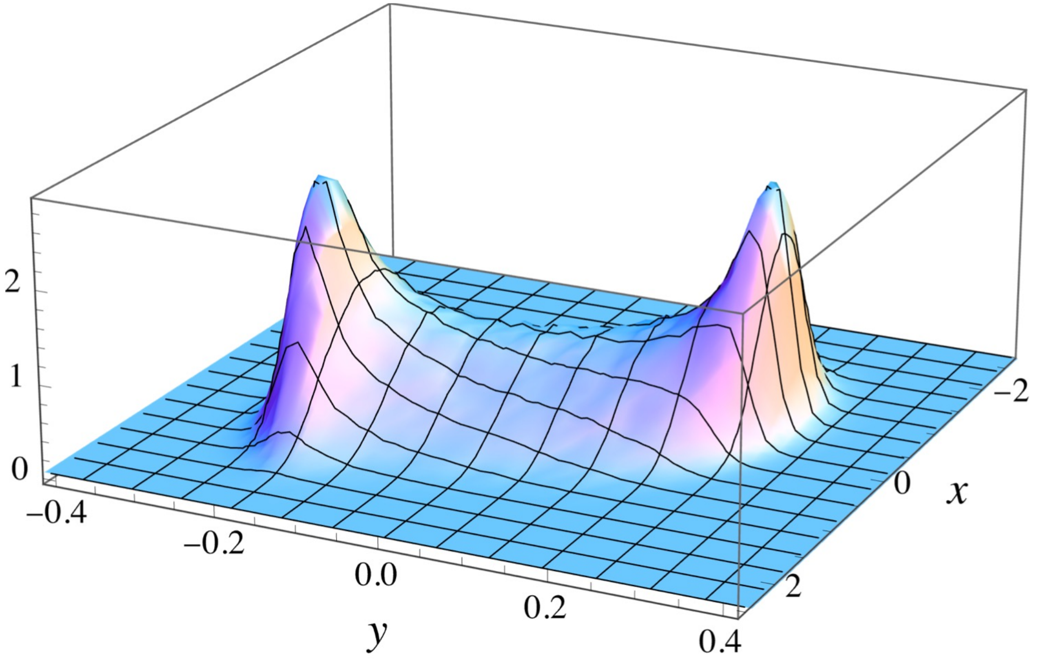
$y_- \approx 0.3029$







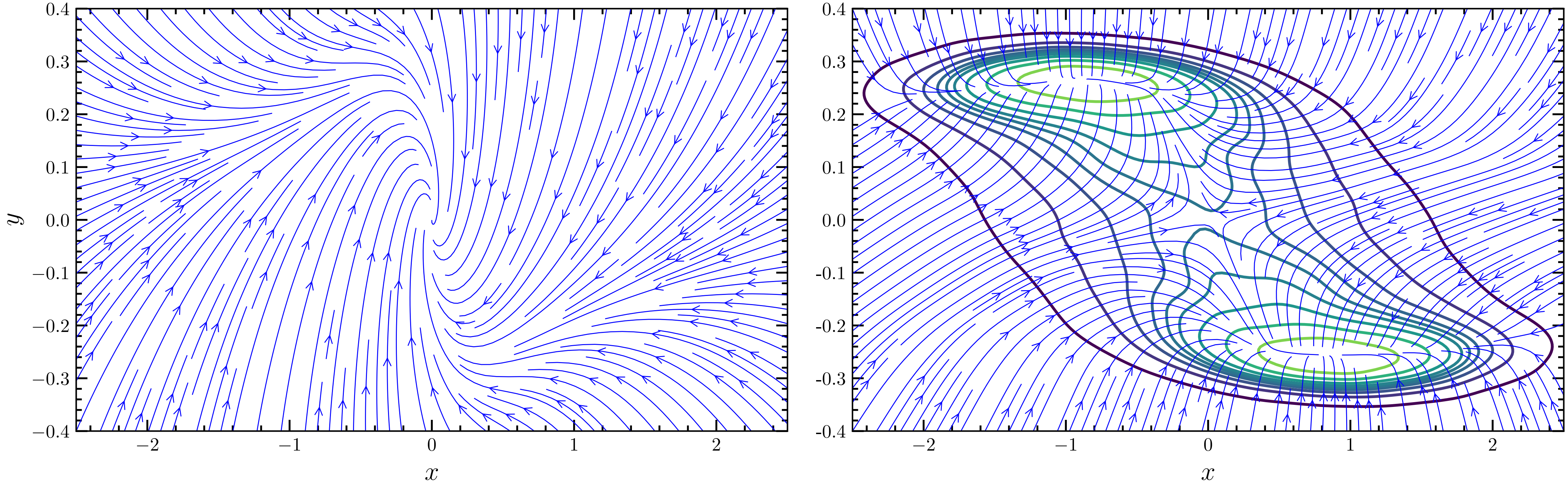
solution of FPE using double expansion in Hermite polynomials



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

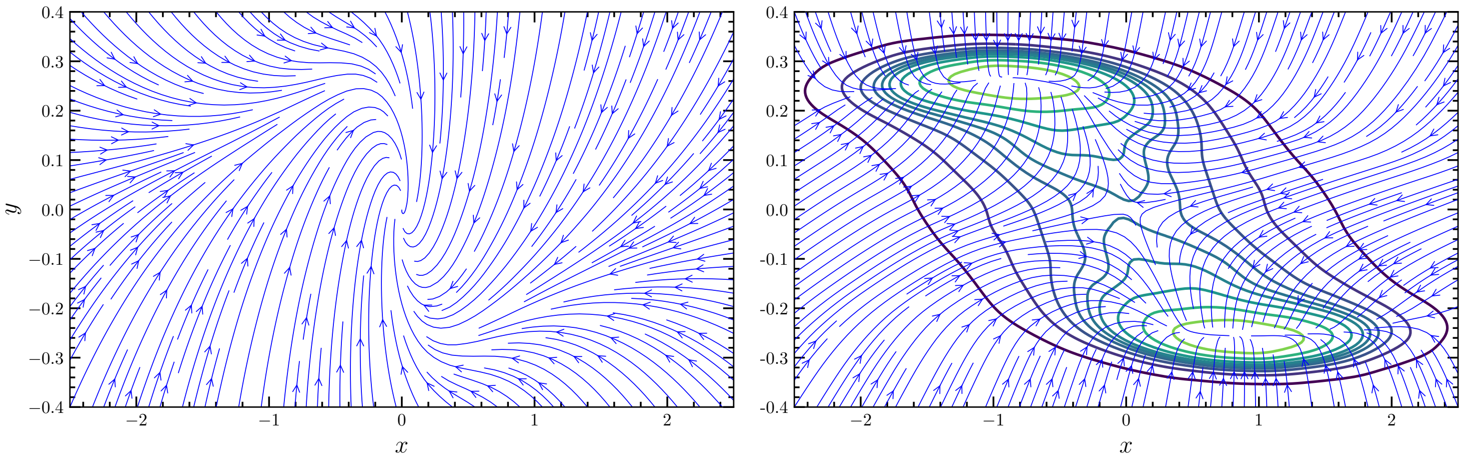
very different processes

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

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
- apply to theories with fermions: DMs learn presence of fermions implicitly?