Higher-order cumulants in diffusion models

Gert Aarts[1], Diaa Eddin Habibi[1], Lingxiao Wang[2], Kai Zhou[3,4]

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

Introduction & Overview

Problem statement: Diffusion models

- Often stated that the data at the end of the noising process is close to random [4].
- How correlations beyond Gaussian ones evolve is less understood.
- Cumulants/Higher-order correlations are of relevance in lattice field theory.

Our contribution [1]

- We derive explicit expressions for the cumulants' evolution for variance-expanding (VE) & variance-preserving (VP) processes.
- In the VE scheme, higher-order cumulants remain constant during evolution.
- In the VP scheme, higher-order cumulants decay exponentially in the noising process.
- We demonstrate these facts numerically for,
  - Double peak toy model (1D data),

$$P(x;\mu,\sigma^2) = \frac{1}{2}\left[\mathcal{N}(x;\mu_0,\sigma_0^2) + \mathcal{N}(x;-\mu_0,\sigma_0^2)\right]$$

  - Lattice scalar field theory (32x32 images) [2]

$$S[\phi] = \sum_x \left[-2\kappa \sum_{\nu=1}^{n} \phi_x\phi_{x+\nu} + \phi_x^2 + \lambda(\phi_x^2-1)^2\right],$$

$$P[\phi] = \frac{1}{\mathcal{Z}} e^{-S[\phi]}, \qquad \mathcal{Z} = \int \mathcal{D}\phi\, e^{-S[\phi]}.$$

Diffusion Models

- Diffusion models learn distributions from data
- Relies on two stochastic processes [4] to learn the score function $\nabla \log P$

$$\partial_t \phi(x,t) = K[\phi(x,t),t] + g(t)\eta(x,t) \quad \text{(Forward)},$$

$$\partial_\tau \phi(x,T-\tau) = -K[\phi(x,\tau),T-\tau] + g^2(T-\tau)\nabla_\phi \log P(\phi,T-\tau) + g(T-\tau)\eta(x,\tau) \quad \text{(Backward)}.$$

- Estimate the score of a distribution based on a sample dataset of the target distribution with score-matching
  - Approximate score function with model $s_\theta(\phi,t)$
  - Train using the Fisher objective [3]

$$L_\theta \propto E_{p_t(x)}[\|s_\theta(\phi,t) - \nabla \log P(\phi,t)\|^2]$$

- Obtain samples using the estimated score, $s_\theta(x,t) \approx \nabla \log P(\phi,t)$, via backwards process
- Use time-conditioned architectures: Feedforward NN (toy model) [1] & U-net (field theory) [2]
- Weighted training objective [4]:

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

where $\lambda(t)$ is chosen to be the variance of the noise at time t.

Generating Functional

For the case of linear drift $K[\phi(x,t),t] = -\frac{1}{2}k(t)\phi(x,t)$, the solution of the forward process:

$$\phi(x,t) = \phi(x,0)f(t,0) + \int_0^t ds\, f(t,s)g(s)\eta(x,s),$$

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

- Moments and cumulants are generated by $Z[J] = \mathbb{E}[e^{J(x)\phi(x,t)}]$ and $W[J] = \log Z[J]$, respectively.
- Higher-order cumulants are then given by

$$\kappa_{n>2}(t) = \left.\frac{\delta^n W[J]}{\delta J(x,t)^n}\right|_{J=0}$$

- First two even-order cumulants read

$$\kappa_2(t) = \mu_2(0)f^2(t,0) + \Xi(t), \qquad \Xi(t) = \int_0^t ds\, f^2(t,s)g^2(s)$$

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

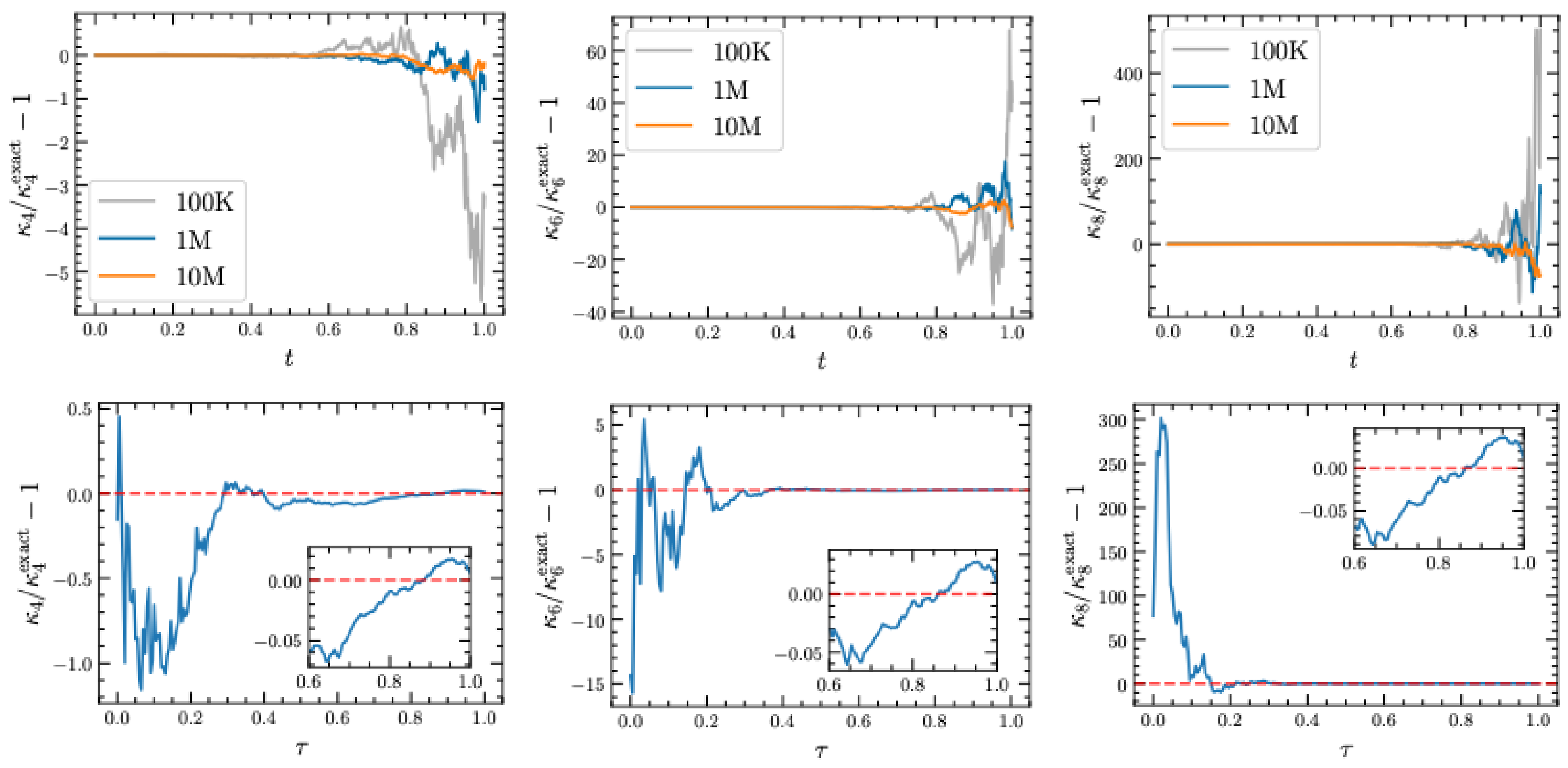
- In general, we find [1]

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

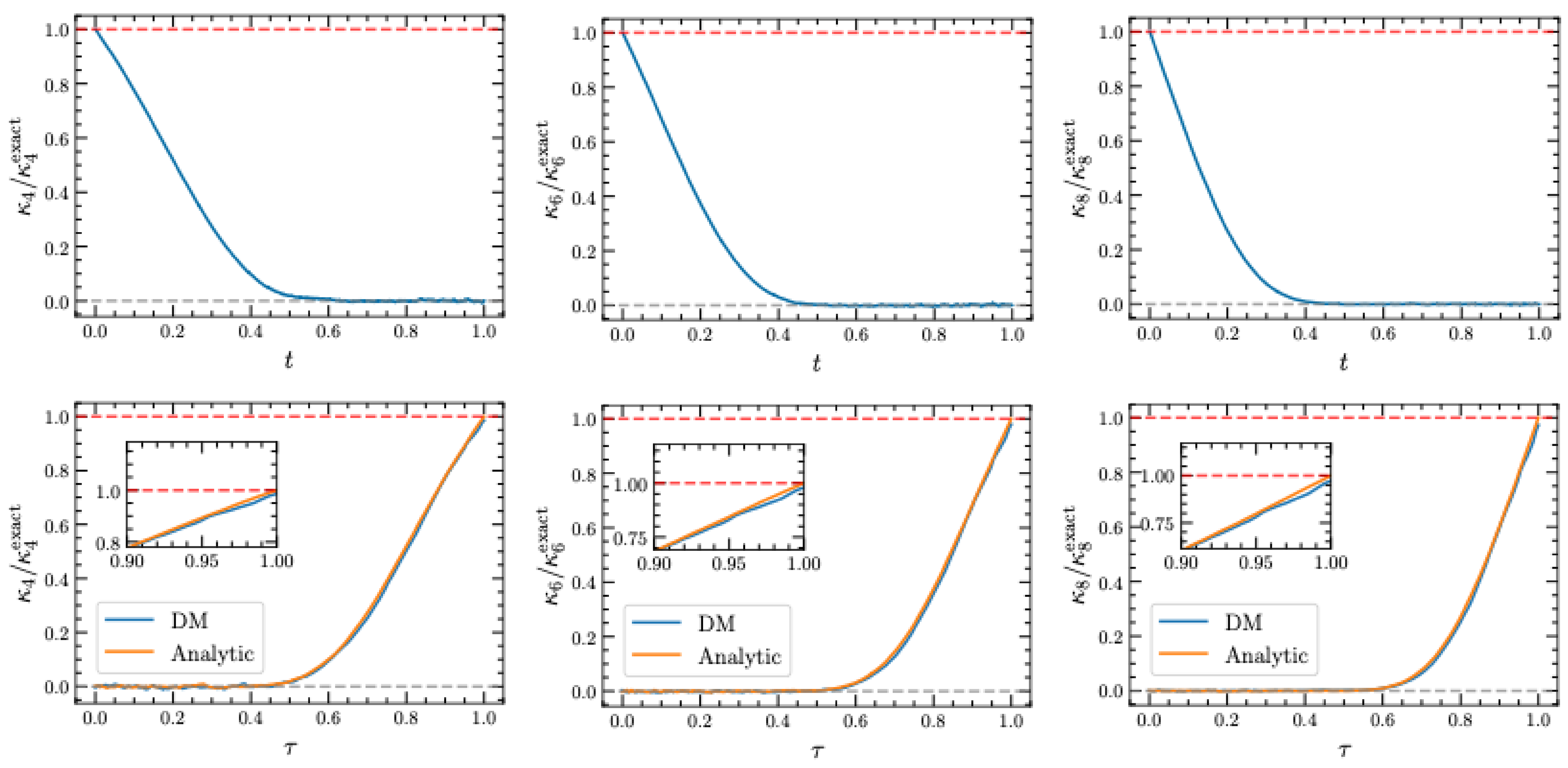
- In pure diffusion (VE), $f(t,s)=1$, cumulants remain constant. Final distribution of the forward process remains as correlated as the target distribution.

Results

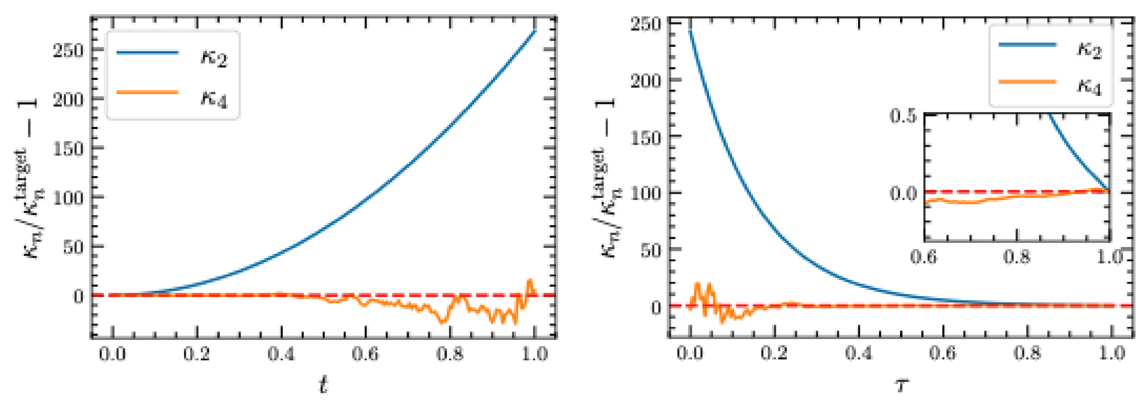
The numerical work shown below supports our analytical predictions. We present the evolution of the normalised 4th, 6th and 8th cumulants in the two-peak model with $\mu_0 = 1$, $\sigma_0 = 1/4$, in the variance-expanding scheme, during the forward process, using 10^5, 10^6 and 10^7 trajectories (above), and during the backward process, with the score determined by the diffusion model, using 10^6 trajectories (below).



Similarly, we show the evolution of the normalised 4th, 6th and 8th cumulants for the same model in the variance-preserving scheme, during the forward process (above), and the backward process (below) using 10^6 trajectories.



Finally, we show the normalised 2nd and 4th in the two-dimensional ϕ^4 field theory, with $\kappa = 0.4, \lambda = 0.022$ and 10^5 configurations on a 32^2 lattice, during the forward (left) and backward (right) process in the VE scheme.



The table shows the first four non-vanishing cumulants κ_n in the scalar ϕ^4 field theory using normalised HMC data and as obtained from the 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)

Outlook

- Investigate alternative noise scheduling to control the statistical behaviour.
- Include density estimations and develop algorithm to be exact via accept/reject step.
- Study the evolution of the non-local correlation function in the diffusion process.
- Apply to large scale lattice simulations.

References

[1] G. Aarts, D.E. Habibi, L. Wang and K. Zhou, *On learning higher-order cumulants in diffusion models*, 2410.21212.
[2] L. Wang, G. Aarts and K. Zhou, *Diffusion models as stochastic quantization in lattice field theory*, JHEP **05** (2024) 060, 2309.17082.
[3] A. Hyvärinen, *Estimation of Non-Normalized Statistical Models by Score Matching*, Journal of Machine Learning Research **6** (2005) 695.
[4] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon and B. Poole, *Score-based generative modeling through stochastic differential equations*, 2011.13456.

Swansea University Prifysgol Abertawe

AIMLAC UKRI Centre for Doctoral Training in Artificial Intelligence, Machine Learning and Advanced Computing

香港中文大學(深圳) The Chinese University of Hong Kong, Shenzhen

iTHEMS