Higher-order cumulants in diffusion models

Gert Aarts and Diaa E. Habibi
Department of Physics, Swansea University, Swansea, SA2 8PP, United Kingdom
{g.aarts, n.e.habibi}@swansea.ac.uk

Lingxiao Wang
Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN
Wako, Saitama 351-0198, Japan
lingxiao.wang@riken.jp

Kai Zhou
School of Science and Engineering, The Chinese University of Hong Kong
Shenzhen (CUHK-Shenzhen), Guangdong, 518172, China
Frankfurt Institute for Advanced Studies, D-60438, Frankfurt am Main, Germany
zhoukai@cuhk.edu.cn

November 22, 2024

Abstract

To analyse how diffusion models learn correlations beyond Gaussian ones, we study the behaviour of higher-order cumulants under both the forward and backward process. We present explicit expressions for the moment- and cumulant-generating functionals, in terms of the distribution of the initial data and properties of the forward process. We show analytically that higher-order cumulants are conserved under pure diffusion, i.e., in models without drift, during the forward process, and that therefore the endpoint of the forward process maintains non-trivial correlations. We demonstrate that since these correlations are encoded in the score function, higher-order cumulants are learnt quickly in the backward process, also when starting from a normal prior. We confirm our analytical results in an exactly solvable toy model and in scalar lattice field theory.

1 Introduction

Diffusion models (DMs) [1] are a widely used class of deep generative models, able to generate high-quality images and videos via a stochastic denoising process (see e.g. Stable Diffusion [2] and DALL-E 2 [3]). In DMs, images are scrambled during the forward process, by applying random noise drawn from a normal distribution to each pixel. It is often stated that at the end of the forward process the images are close to being fully random, that is, without any correlations remaining [4, 5]. During the forward process, the change in the logarithm of the distribution function is learned ("score matching") [6]. In the backward process, this score is applied to initial conditions drawn from a normal distribution and new images are generated ("denoising") [7].

To understand how correlations beyond Gaussian ones evolve during both the forward and backward process, we use lattice field theory (LFT) as a robust and well-understood framework to address this question. Indeed, in field theory, interactions between fundamental fields are encoded in higher n-point correlation functions and there is a long history [8] of studying strongly interacting quantum

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

field theories numerically by combining the path integral formulation with Monte Carlo methods, after discretisation on a spacetime lattice [9, 10].

Besides using field theory as a framework to analyse the dynamics of DMs, machine learning (ML) methods also offer a new avenue for generating LFT configurations, with field configurations playing the role of images, potentially avoiding issues related to critical slowing down (see e.g., Refs. [11, 12]). The first applications of DMs to LFT can be found in Refs. [13, 14]. To apply DMs to LFT, it is of paramount importance to understand whether higher-order correlations are faithfully reproduced. In this contribution we study the dynamics of cumulants to address both questions. More details of the work presented here can be found in Ref. [15].

Related work Applications of field theory to understand ML models can be found in e.g. Refs. [16–18]. The application of generative ML methods to LFT is reviewed in Ref. [19]. Flow-based methods were developed in Refs. [20, 21] and subsequent work. Diffusion models were first applied to LFT in Refs. [13, 22], in which the connection with stochastic quantisation [23–25] was pointed out. A formulation using Feynman's path integral was given in Ref. [26]. See Ref. [15] for more references.

2 Diffusion models, moments and cumulants

We consider a real scalar field $\phi(x)$, with the target probability distribution

$$P_0[\phi] = \frac{1}{Z} e^{-S[\phi]}, \qquad Z = \int D\phi\, e^{-S[\phi]}, \tag{1}$$

where $S[\phi]$ is the Euclidean action. The integral is over all field configurations and $D\phi$ denotes the path integral measure. Below we assume that the first moment vanishes, or has been subtracted, $\phi(x) \to \phi(x) - \mathbb{E}_{P_0}[\phi(x)]$. This distribution is used to generate initial configurations for the forward process, given by

$$\partial_t \phi(x,t) = K[\phi(x,t), t] + g(t)\eta(x,t). \tag{2}$$

Here we use the DM formulation in terms of stochastic differential equations (SDEs) [4], with $0 \le t \le T$, $K[\phi, t]$ a possible drift term, $g(t)$ the strength of the noise, and $\eta \sim \mathcal{N}(0,1)$. Below we consider both score-based, variance-expanding schemes [7, 27], in which $K[\phi, t] = 0$, and denoising diffusion probabilistic models (DDPMs) [28, 29], in which $K[\phi, t] = -\frac{1}{2} g^2(t)\phi(x,t)$. We take $g(t) = \sigma^{t/T}$, with σ a tunable but generically large parameter.

The corresponding backward process reads

$$\partial_\tau \phi(x,\tau) = -K[\phi(x,\tau), T-\tau] + g^2(T-\tau)\nabla_\phi \log P(\phi, T-\tau) + g(T-\tau)\eta(x,\tau), \tag{3}$$

where $\tau = T - t$. Initial conditions for the backward process are drawn from a normal distribution with a variance comparable to the final variance of the forward process. The second term in Eq. (3) is the "score", the change in the logarithm of the distribution, which is determined during the forward process, via score matching [6].

With a linear drift, $K[\phi(x,t), t] = -\frac{1}{2} k(t)\phi(x,t)$, it is straightforward to solve Eq. (2) analytically, as

$$\phi(x,t) = \phi_0(x) f(t,0) + \int_0^t ds\, f(t,s) g(s) \eta(x,s), \qquad f(t,s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}, \tag{4}$$

where the initial condition $\phi_0 \sim P_0[\phi_0]$. Note that for pure diffusion, with $k(t) = 0$, $f(t,s) = 1$.

We consider moments, involving powers of the field at coinciding spacetime points,

$$\mu_n(x,t) = \mathbb{E}[\phi^n(x,t)]. \tag{5}$$

Cumulants $\kappa_n(x,t)$ are obtained via the relation [30]

$$\kappa_n = \mu_n - \sum_{m=2}^{n-2} \binom{n-1}{m-1} \kappa_m \mu_{n-m}, \tag{6}$$

with $\mu_1 = \kappa_1 = 0$. When the target theory is translationally invariant, moments and cumulants are independent of x and we drop the x-label. The time-dependent second moment (or second cumulant) then reads

$$\mu_2(t) = \kappa_2(t) \equiv \mathbb{E}[\phi^2(x,t)] = \mathbb{E}_{P_0}[\phi_0^2(x)] f^2(t,0) + \Xi(t) = \mu_2(0) f^2(t,0) + \Xi(t), \tag{7}$$

where $\mu_2(0)$ is the second moment of the target distribution and

$$\Xi(t) = \int_0^t ds\, f^2(t,s) g^2(s). \tag{8}$$

The fourth moment and cumulant are given by, after some algebra,

$$\mu_4(t) = \mu_4(0) f^4(t,0) + 6\mu_2(0) f^2(t,0)\Xi(t) + 3\Xi^2(t), \tag{9}$$

$$\kappa_4(t) = \mu_4(t) - 3\mu_2^2(t) = \left[\mu_4(0) - 3\mu_2^2(0)\right] f^4(t,0) = \kappa_4(0) f^4(t,0), \tag{10}$$

i.e. the fourth cumulant is equal to the fourth cumulant of the target theory times a time-dependent function, which is 1 for pure diffusion. This structure turns out to be the same for all higher-order cumulants, which can be shown using generating functionals [15]. Moments are generated by

$$Z[J] = \mathbb{E}[e^{J(x,t)\phi(x,t)}] = e^{\frac{1}{2}J^2(x,t)\Xi(t)} \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x) f(t,0)}, \tag{11}$$

and the cumulant-generating functional reads

$$W[J] = \log Z[J] = \frac{1}{2} J^2(x,t)\Xi(t) + \log \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x) f(t,0)}. \tag{12}$$

The higher-order cumulants are then given by

$$\kappa_{n>2}(t) = \frac{\delta^n W[J]}{\delta J(x,t)^n}\Big|_{J=0} = \frac{\delta^n}{\delta J(x,t)^n} \log \mathbb{E}_{P_0}[e^{J(x,t)\phi_0(x) f(t,0)}]\Big|_{J=0}, \tag{13}$$

and hence equal to the cumulants in the target theory, multiplied with the time-dependent function $f^n(t,0)$. In particular, for pure diffusion we find $\kappa_{n>2}(t) = \kappa_n(0)$. We conclude that in that case the final distribution of the forward process is not a simple normal distribution, but is as correlated as the target distribution. These higher-order cumulants are encoded in the score and should be reconstructed during the backward process, which we explore now.

3 Applications

In this section we verify our analytical results in two cases, an exactly solvable toy model with one degree of freedom and a self-interacting scalar field theory in two dimensions. In the toy model we train a DM using a time-conditioned fully connected neural network and in the scalar field theory we use a U-Net architecture. More details on the implementation can be found in Refs. [13, 15].

3.1 Exactly solvable toy model

We first consider as target distribution a linear combination of two normal distributions for one degree of freedom,

$$P_0(x) = \frac{1}{2}\left[\mathcal{N}(x;\mu,\sigma_0^2) + \mathcal{N}(x;-\mu,\sigma_0^2)\right]. \tag{14}$$

All odd moments and cumulants vanish. All even moments and cumulants can be computed exactly and are nonzero. We have solved this model numerically, using a DM in both a variance-expanding and a variance-preserving (DDPM) scheme, with $g(t) = \sigma^{t/T}$, and $\sigma = 10$ and $T = 1$.

We show the evolution of the $4^{\text{th}}, 6^{\text{th}}$ and 8^{th} cumulants for the variance-expanding scheme in Fig. 1 and for the DDPM in Fig. 2, during the forward (top row) and the backward (bottom row) processes, for the parameter choice $\mu_0 = 1$, $\sigma_0 = 1/4$. The cumulants behave as predicted analytically. In the variance-expanding case, they are approximately constant, except towards the end (start) of the forward (backward) process. This is due to the need for precise cancellations, which requires sufficient statistics. For the forward process, this is demonstrated by including expectation values with $10^5, 10^6$ and 10^7 trajectories. The apparent numerical instability is reduced as the number of trajectories increases. This supports the analytical result that the higher-order cumulants are preserved. In the backward process we observe the reverse behaviour, with only a partial cancellation initially. After some time, the cumulants become approximately constant and equal to the target value, see Table 1. In the DDPM case, the higher-order cumulants evolve from the target value to zero, and vice versa, in a much more controlled manner than in the variance-expanding scheme; the cancellations required above are not needed here. We also show the result obtained using the analytically determined score [15]. A comparison between the DM results and the expected ones is given in Table 1. Overall we conclude that the higher-order cumulants are learned correctly, with the variance-expanding scheme slightly outperforming the DDPM.

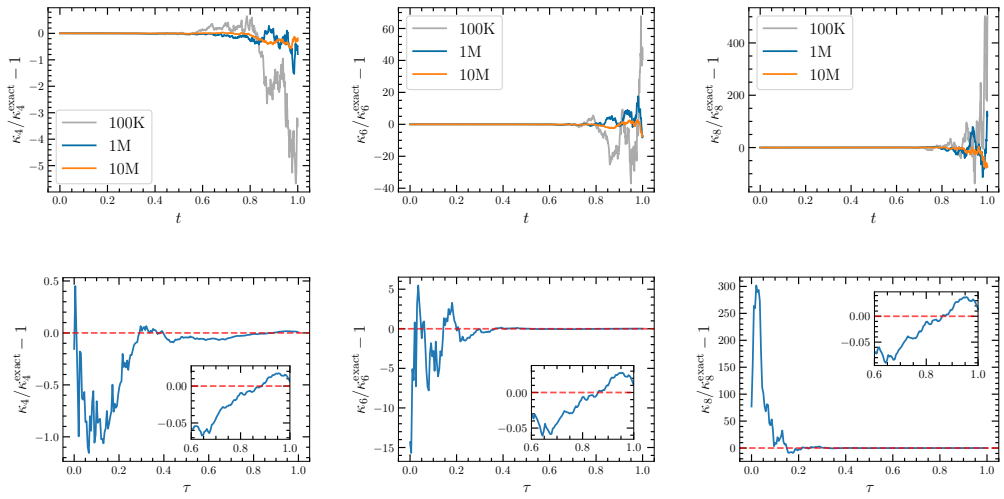

Figure 1: Evolution of the normalised 4^{th}, 6^{th} and 8^{th} cumulants, presented as $\kappa_n/\kappa_n^{\text{exact}} - 1$, in the two-peak model with $\mu_0 = 1$ and $\sigma_0 = 1/4$, in the variance-expanding scheme, during the forward process, using 10^5, 10^6 and 10^7 trajectories (above), and during the backward process, with the score determined by the diffusion model, using 10^6 trajectories (below).

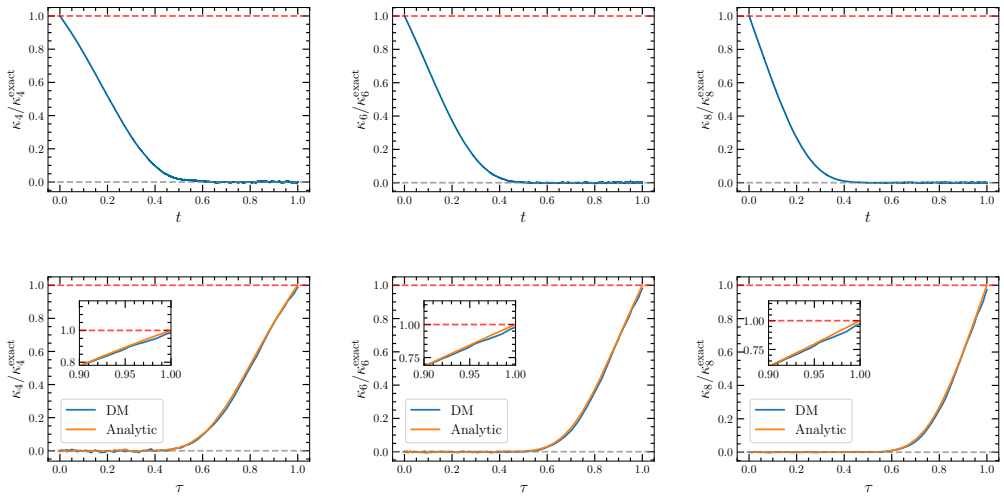

Figure 2: As above, using the DDPM, presented as $\kappa_n/\kappa_n^{\text{exact}}$, with 10^6 trajectories. In the backward process, both the score determined by the diffusion model and the analytical score are used.

3.2 Self-interacting scalar field

We now extend the analysis to a scalar field $\phi(x)$, defined on a two-dimensional lattice, with a $\lambda\phi^4$ interaction. We follow the notation of Ref. [13]. The results shown below are obtained on a 32×32 lattice, with hopping parameter $\kappa = 0.4$ (not to be confused with a cumulant) and coupling $\lambda = 0.022$. We used 10^5 configurations to train the model using the variance-expanding scheme with $\sigma = 25$, and also 10^5 configurations to evolve the cumulants during the forward and backward processes.

Fig. 3 shows the 2^{nd}, 4^{th} and 6^{th} cumulant, normalised with the numerically computed target value, during the forward (left) and backward (right) process. As expected, the 2^{nd} moment or cumulant increases as $\sigma^{2t/T}$ (decreases as $\sigma^{2(T-\tau)/T}$). The 4^{th} and and 6^{th} cumulants are approximately constant, with again finite-statistics effects towards the end (start) of the forward (backward) process,

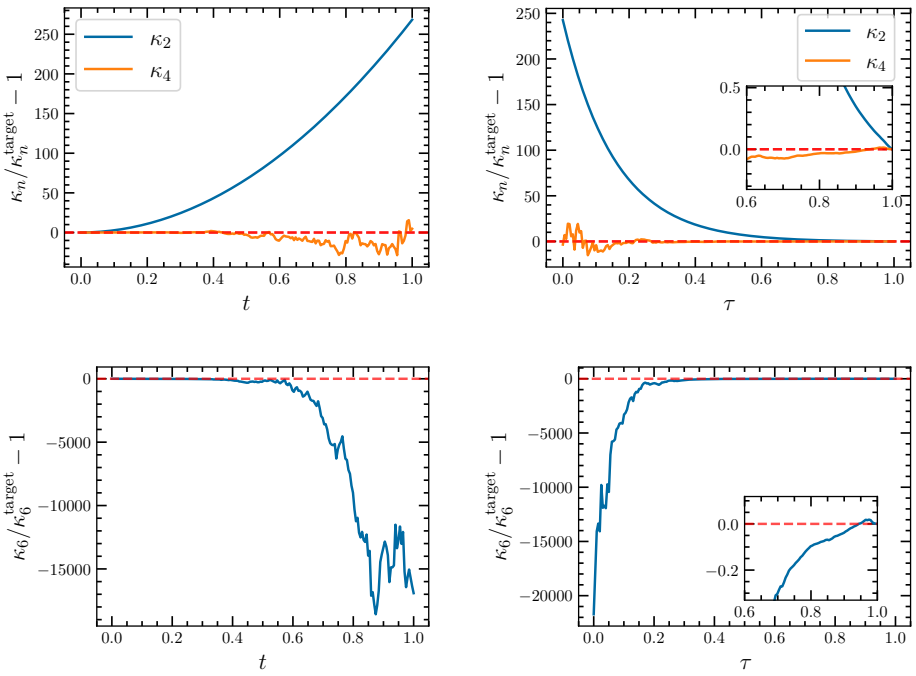

Figure 3: Evolution of the normalised 2^{nd} and 4^{th} (above) and 6^{th} (below) cumulant, presented as $\kappa_n/\kappa_n^{\text{target}} - 1$, in the two-dimensional ϕ^4 theory, during the forward (left) and backward (right) process in the variance-expanding scheme.

as above. Importantly, the cumulants are reproduced at the end of the backward process, as shown in the insets and in Table 1.

Two-peak model	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)	$-265.6(6)$
Scalar field theory	κ_2	κ_4	κ_6	κ_8
HMC (normalised)	0.39597(4)	$-0.29453(6)$	0.90108(28)	$-5.8689(25)$
Variance expanding	0.39598(4)	$-0.29454(7)$	0.90113(32)	$-5.8694(28)$

Table 1: First four nonvanishing cumulants κ_n in the two-peak model, as obtained from training data and from diffusion models using 10^6 configurations, and in the scalar ϕ^4 field theory, with $\kappa = 0.4$, $\lambda = 0.022$ and 10^5 configurations on a 32^2 lattice, using normalised HMC data and as obtained from the diffusion model. Statistical errors are computed by bootstrapping in both cases.

4 Summary

Employing and motivated by lattice field theory, we have studied the evolution of higher-order cumulants during the forward and backward process in diffusion models, using both variance-expanding and preserving schemes. We have shown analytically that cumulants κ_n, with $n > 2$, are preserved in the case of pure diffusion, i.e. in the absence of a drift term. Hence the final configurations of the forward process remain highly correlated. These correlations are learned during the backward process, as they are encoded in the score. Our analytical findings are supported by numerical evidence in an exactly solvable toy model and in the case of a self-interacting scalar field defined on a two-dimensional lattice.

Acknowledgements – Part of the work presented here was carried out at the ECT* workshop *Machine Learning and the Renormalisation Group* in May 2024. We thank the participants of this meeting for discussion and ECT* for support. GA is supported by STFC Consolidated Grant ST/T000813/1. DEH is supported by the UKRI AIMLAC CDT EP/S023992/1. LW thanks the DEEP-IN working group at RIKEN-iTHEMS for its support in the preparation of this paper. KZ is supported by the CUHK-Shenzhen university development fund under grant No. UDF01003041 and UDF03003041, and Shenzhen Peacock fund under No. 2023TC0179.

We acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government.

Research Data and Code Access – The code and data used for this manuscript are available from Ref. [31].

References

[1] L. Yang, Z. Zhang, S. Hong, R. Xu, Y. Zhao, Y. Shao et al., *Diffusion Models: A Comprehensive Survey of Methods and Applications*, 2209.00796.

[2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, *High-Resolution Image Synthesis with Latent Diffusion Models*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022) 10684 [2112.10752].

[3] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu and M. Chen, *Hierarchical Text-Conditional Image Generation with CLIP Latents*, *arXiv e-prints* (2022) [2204.06125].

[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.

[5] P. Nakkiran, A. Bradley, H. Zhou and M. Advani, *Step-by-Step Diffusion: An Elementary Tutorial*, 2406.08929.

[6] A. Hyvärinen, *Estimation of Non-Normalized Statistical Models by Score Matching*, *Journal of Machine Learning Research* **6** (2005) 695.

[7] Y. Song and S. Ermon, *Generative Modeling by Estimating Gradients of the Data Distribution*, 1907.05600.

[8] M. Creutz, *Monte Carlo Study of Quantized SU(2) Gauge Theory*, *Phys. Rev. D* **21** (1980) 2308.

[9] C. Gattringer and C.B. Lang, *Quantum chromodynamics on the lattice*, vol. 788, Springer, Berlin (2010), 10.1007/978-3-642-01850-3.

[10] J. Smit, *Introduction to quantum fields on a lattice: A robust mate*, vol. 15, Cambridge University Press (1, 2011).

[11] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.

[12] ALPHA collaboration, *Critical slowing down and error analysis in lattice QCD simulations*, *Nucl. Phys. B* **845** (2011) 93 [1009.5228].

[13] L. Wang, G. Aarts and K. Zhou, *Diffusion models as stochastic quantization in lattice field theory*, *JHEP* **05** (2024) 060 [2309.17082].

[14] Q. Zhu, G. Aarts, W. Wang, K. Zhou and L. Wang, *Diffusion models for lattice gauge field simulations*, in *38th conference on Neural Information Processing Systems*, 2024 [2410.19602].

[15] G. Aarts, D.E. Habibi, L. Wang and K. Zhou, *On learning higher-order cumulants in diffusion models*, 2410.21212.

[16] J. Halverson, A. Maiti and K. Stoner, *Neural Networks and Quantum Field Theory*, *Mach. Learn. Sci. Tech.* **2** (2021) 035002 [2008.08601].

[17] M. Demirtas, J. Halverson, A. Maiti, M.D. Schwartz and K. Stoner, *Neural network field theories: non-Gaussianity, actions, and locality*, *Mach. Learn. Sci. Tech.* **5** (2024) 015002 [2307.03223].

[18] G. Aarts, B. Lucini and C. Park, *Scalar field restricted Boltzmann machine as an ultraviolet regulator*, *Phys. Rev. D* **109** (2024) 034521 [2309.15002].

[19] K. Cranmer, G. Kanwar, S. Racanière, D.J. Rezende and P.E. Shanahan, *Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics*, *Nature Rev. Phys.* **5** (2023) 526 [2309.01156].

[20] M.S. Albergo, G. Kanwar and P.E. Shanahan, *Flow-based generative models for Markov chain Monte Carlo in lattice field theory*, *Phys. Rev. D* **100** (2019) 034515 [1904.12072].

[21] G. Kanwar, M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, S. Racanière et al., *Equivariant Flow-Based Sampling for Lattice Gauge Theory*, *Phys. Rev. Lett.* **125** (2020) 121601 [2003.06413].

[22] L. Wang, G. Aarts and K. Zhou, *Generative Diffusion Models for Lattice Field Theory*, in *37th Conference on Neural Information Processing Systems*, 2023 [2311.03578].

[23] G. Parisi and Y.S. Wu, *Perturbation theory without gauge fixing*, *Sci. China, A* **24** (1980) 483.

[24] P.H. Damgaard and H. Hüffel, *Stochastic quantization*, *Phys. Rept.* **152** (1987) 227.

[25] M. Namiki, *Basic ideas of stochastic quantization*, *Prog. Theor. Phys. Suppl.* **111** (1993) 1.

[26] Y. Hirono, A. Tanaka and K. Fukushima, *Understanding Diffusion Models by Feynman's Path Integral*, 2403.11262.

[27] Y. Song and S. Ermon, *Improved techniques for training score-based generative models*, 2006.09011.

[28] J. Ho, A. Jain and P. Abbeel, *Denoising diffusion probabilistic models*, in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, NIPS'20, (Red Hook, NY, USA), pp. 6840–6851, Curran Associates Inc., Dec., 2020.

[29] J. Sohl-Dickstein, E.A. Weiss, N. Maheswaranathan and S. Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn. - Vol. 37*, ICML'15, (Lille, France), pp. 2256–2265, JMLR.org, July, 2015.

[30] Peter J. Smith, *A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa*, *The American Statistician* **49** (1995) 217.

[31] D.E. Habibi, G. Aarts, L. Wang and K. Zhou, *DiaaEddinH/On-learning-higher-order-cumulants-in- diffusion-models: v1.0.2*, 2024. 10.5281/zenodo.14041604.