Machine learning stochastic differential equations for the evolution of order parameters of classical many-body systems in and out of equilibrium

To cite this article: Francesco Carnazza *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 045002

MACHINE
LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
6 June 2024

REVISED
10 September 2024

ACCEPTED FOR PUBLICATION
13 September 2024

PUBLISHED
7 October 2024

PAPER

Machine learning stochastic differential equations for the evolution of order parameters of classical many-body systems in and out of equilibrium

Francesco Carnazza[1,*], Federico Carollo[1], Sabine Andergassen[2], Georg Martius[3,4], Miriam Klopotek[5] and Igor Lesanovsky[1,6]

[1] Institut für Theoretische Physik and Center for Quantum Science, Universität Tübingen, Auf der Morgenstelle 14, 72076 Tübingen, Germany
[2] Institute of Information Systems Engineering and Institute for Solid State Physics, Vienna University of Technology, 1040 Vienna, Austria
[3] Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4, 72076 Tübingen, Germany
[4] Wilhelm Schickard Institut für Informatik, Universität Tübingen, Maria-von-Linden-Straße 6, 72076 Tübingen, Germany
[5] Stuttgart Center for Simulation Science, SimTech Cluster of Excellence EXC 2075, University of Stuttgart, Universitätsstraße 32, 70569 Stuttgart, Germany
[6] Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems, University of Nottingham, Nottingham NG7 2RD, United Kingdom
[*] Author to whom any correspondence should be addressed.

E-mail: francesco.carnazza@uni-tuebingen.de

Keywords: stochastic processes, dynamical Ising model, contact process, machine learning

Abstract

We develop a machine learning algorithm to infer the emergent stochastic equation governing the evolution of an order parameter of a many-body system. We train our neural network to independently learn the directed force acting on the order parameter as well as an effective diffusive noise. We illustrate our approach using the classical Ising model endowed with Glauber dynamics, and the contact process as test cases. For both models, which represent paradigmatic equilibrium and nonequilibrium scenarios, the directed force and noise can be efficiently inferred. The directed force term of the Ising model allows us to reconstruct an effective potential for the order parameter which develops the characteristic double-well shape below the critical temperature. Despite its genuine nonequilibrium nature, such an effective potential can also be obtained for the contact process and its shape signals a phase transition into an absorbing state. Also, in contrast to the equilibrium Ising model, the presence of an absorbing state renders the noise term dependent on the value of the order parameter itself.

1. Introduction

Stochastic processes are fundamentally important in physics [1–3]. For instance, random microscopic fluctuations can strongly impact the evolution of macroscopic physical observables, e.g. order parameters close to phase transitions. Monte Carlo methods [4–6] are often the 'benchmark' for the computational treatment of classical many-body dynamics, allowing for efficient sampling of stochastic microscopic configurations and trajectories. The large-scale dynamics of the order parameter are instead typically modeled by a stochastic differential equation. The latter contains both a force term, leading to a deterministic drift, and a noise term yielding diffusive behavior[7]. [7, 8] However, establishing a connection between fluctuating microscopic stochastic trajectories and the coarse-grained evolution of the order parameter is a challenging task that can rarely be accomplished analytically.

[7] The drift and the diffusion represent the most basic ingredients for a coarse-grained dynamics. More general forms might include memory kernels or other non-Markovian time dependencies [7, 8].

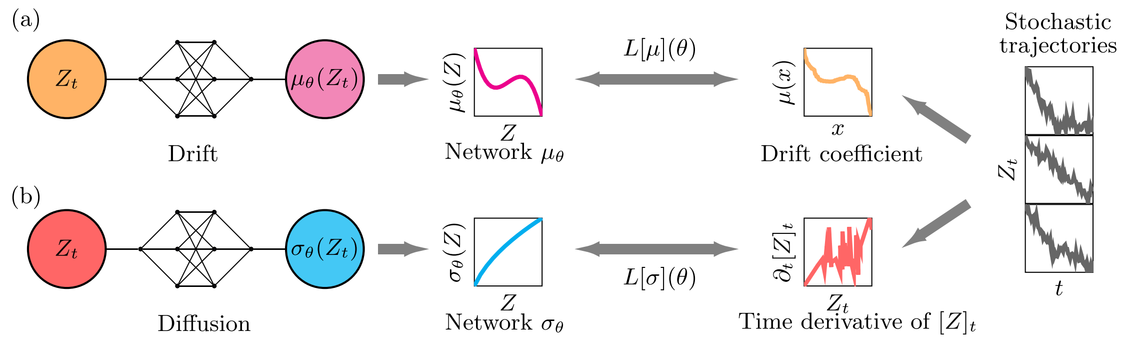

Figure 1. Training of the drift coefficient μ_θ and the diffusion coefficient σ_θ from the stochastic trajectories of the order parameter. (a) The network for the drift coefficient, μ_θ, is trained by minimizing a distance, $L[\mu](\theta)$ in equation (4), between $\mu_\theta(x)$ and the drift coefficient μ, cf equation (3). For a given value x of the order parameter, we consider all the values Z_t close to it in the observed trajectories, and their update Z_{t+dt}. The finite difference $Z_{t+dt} - Z_t$ is then used to locally approximate the drift coefficient μ. (b) The network for the diffusion coefficient σ_θ is trained by minimizing a distance, cf equation (10), between $\sigma_\theta^2(Z_t)$ and the derivative of the quadratic variation $\partial_t[Z]_t$, see $L[\sigma](\theta)$ in equation (9). The quadratic variation $[Z]_t$ itself, cf equation (6), is locally approximated from the observed trajectories.

In this paper, we develop a machine learning approach [9–14] to bridge this gap. To illustrate our method, we consider two paradigmatic classical many-body systems: the 2D Ising model evolving under Glauber dynamics [15–17] and the nonequilibrium contact process in 1D. The dynamics considered for the Ising model obey detailed balance, which eventually takes the system to a state of thermal equilibrium. As a function of temperature, this state shows a transition from a paramagnetic to a ferromagnetic state, characterized by a zero and non-zero value of the order parameter, respectively. As we will show, this transition manifests in the structure of the learned drift term cf figure 1(a), from which one can reconstruct an effective potential that exhibits a characteristic double-well shape below the critical temperature. Both the paramagnetic and ferromagnetic phases are fluctuating, which is also reflected in the learned noise term. In contrast to the scenario of the Ising model, the contact process represents a genuine out-of-equilibrium system [18–21], i.e. its dynamics does not obey detailed balance and its stationary state is non-thermal. The model features a phase transition between a non-fluctuating absorbing state in which the order parameter is strictly zero and a fluctuating active phase with a non-vanishing order parameter. Interestingly, we show that also for this genuine nonequilibrium process, an effective potential governing the deterministic drift of the order parameter can be constructed using our machine learning approach. Unlike for the Ising model, however, where the learned noise is such that both phases are fluctuating, a noise term is inferred whose strength depends on the order parameter. In particular, the noise strength tends to zero for vanishing values of the order parameter, see sketch in figure 1(b), signalling an approach to the (non-fluctuating) absorbing state.

Our method is applicable to a wide range of many-body processes in and out of equilibrium. It provides a way to determine a stochastic equation for order parameters which is intuitive and directly interpretable, as in mean-field theories. Remarkably, it also carries information about the exact low-dimensional physics of the considered model, as we demonstrate through estimates of critical exponents. Moreover, our method should also be applicable for inferring effective stochastic differential equations for the evolution of order parameters from experimental data. Our approach is based on learning ordinary differential equations, which are relatively straightforward to handle. In contrast, alternative methods focus on learning the full probability distribution governing the stochastic process [22–27]. These approaches often require assumptions about the functional form of the probability distribution, such as modeling it with a set of Gaussian distributions. Other techniques learn the stochastic differential equation by integrating the stochastic dynamics and optimizing over the probability distribution of the variable [28–34], but they encounter difficulties in the scenarios we focus on, where the data is subject to significant inherent noise. As background, we explain more in detail what are the physical quantities that we examine. These quantities describe macroscopic properties of the microscopic spin models, which are referred to as ”order parameters” in statistical mechanics. Order parameters, as discussed in [35], are named for their role in measuring the degree of order within a physical system. Typically, they have non-zero values in phases where some form of order exists, and zero values in disordered phases. The coarse-grained quantities we focus on are the averages of the binary spin values over the lattice size at each time step. We assume that their dynamics can be described by a stochastic differential equation of the Itô type whose drift and diffusion coefficient are time-independent. In figure 2, we present various configurations of the microscopic spin system for the Ising model. Specifically, in figure 2(a), we illustrate several configurations of the microscopic spin system and the

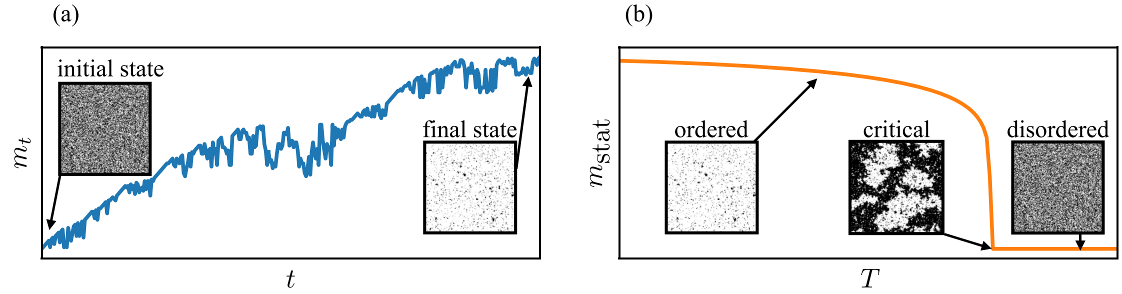

Figure 2. The stochastic dynamics of the order parameter and its steady state for the dynamical Ising model. In panel (a), the evolution of the order parameter, specifically the magnetization m_t, for the dynamical Ising model is depicted as a function of time t at a specific value of the control parameter, namely, the temperature T. The insets show the actual corresponding configurations of the entire spin system at the initial and final states. In the Ising model, varying the temperature results in three distinct phases: the ordered phase, the critical phase, and the disordered phase. The critical phase appears at the critical temperature T_c (see panel (b)). Below the critical temperature, the configurations evolve towards a stationary magnetization per single spin m_{stat} that is non-zero.

corresponding time evolution of the order parameter. In figure 2(b), we display the stationary values of the states as a function of the control parameter, which in this case is the temperature.

2. Many-body stochastic processes

2.1. The evolution of stochastic observables

For the sake of concreteness, we focus on many-body lattice systems of N sites, each of which is associated with a classical spin variable. We denote the system state, or system configuration, through the vector s containing the values s^i of the variables at the different sites i. We furthermore assume the system to be subject to a discrete-time Markovian stochastic spin-flip dynamics.

Relevant information about the above many-body system is provided by so-called order parameters, which encode properties of the whole configuration. A paradigmatic example is given by an average of the form $Z_t \equiv Z(s_t) = \frac{1}{N}\sum_i s_t^i$, where s_t is the time-evolving state of the system. As a consequence of the stochastic nature of s_t, also the effective dynamics of Z_t is stochastic. For large systems and at a continuous coarse-grained time scale, Z_t becomes a continuous random variable that may be expected to obey an emergent stochastic differential equation of the form

$$\mathrm{d}Z_t = \mu\left(Z_t, t\right)\mathrm{d}t + \sigma\left(Z_t, t\right)\mathrm{d}W_t\,. \tag{1}$$

Here, the function μ is referred to as the drift coefficient, while σ is called diffusion coefficient. W_t is a standard Wiener process [1] and $\mathrm{d}W_t$ is its increment satisfying the relations $\mathbb{E}[\mathrm{d}W_t] = 0$ and $\mathbb{E}[\mathrm{d}W_t^2] = \mathrm{d}t$, with $\mathbb{E}$ denoting expectation over the noise. Despite the simple form of equation (1), understanding the functional form of μ and σ is in general a difficult task. In the following, we propose a method to learn an approximation to the analytical form of the drift and the diffusion coefficients by means of neural networks. We determine two artificial neural networks μ_θ and σ_θ (see sketch in figure 1), which describe the dynamics of Z_t, given the network parameters (weights and biases) θ. We restrict ourselves to the Markovian case in which μ_θ and σ_θ do not depend on time

$$\mathrm{d}Z_t = \mu_\theta\left(Z_t\right)\mathrm{d}t + \sigma_\theta\left(Z_t\right)\mathrm{d}W_t. \tag{2}$$

To approximate the functions μ and σ we use a data-driven method, i.e. the networks μ_θ and σ_θ are trained on a data set composed of trajectories Z_t, which we call ground truth data, see also figure 1. Note that restricting to the Markovian case of equation (2) is an assumption since, even if the dynamics of the system configuration s_t is Markovian at the microscopic scale, the emergent dynamics of the order parameters—i.e. macroscopic quantities—may feature non-Markovian effects.

2.2. Neural network representation of the drift and diffusion coefficients

Our approach consists of training the networks μ_θ and σ_θ with separate routines, independently from each other.

As we discuss below, this means that the data sets used for training, despite being derived from the same sets of trajectories, differ from each other. Specifically, in the case of the drift coefficient, the network is trained using the first moments of the (infinitesimal) finite differences between adjacent time steps in the trajectories. For the diffusion coefficient, the network is trained using the time derivative of the quadratic variation, which is akin to the second moment of the finite differences between two different times in the trajectories. In fact, nothing prevents us from training the drift and diffusion terms using completely different sets of trajectories.

The drift term can be quantified by exploiting averages over trajectories of the infinitesimal increment $\mathrm{d}Z_t$ in equation (1). More precisely, starting from equation (1) it is possible to show that the function μ at point x can be obtained as the limit [36–38]

$$\mu\left(x\right) = \lim_{\mathrm{d}t\to 0+} \frac{\mathbf{E}_{Z_t=x}\left[Z_{t+\mathrm{d}t}\right] - x}{\mathrm{d}t}, \tag{3}$$

where $\mathbf{E}_{Z_t=x}$ denotes expectation conditional on the process being in x at time t. In the theory of stochastic processes, the above limit also provides the action of the so-called infinitesimal generator $\mathcal{W}$ on the function x, $\mu(x) = \mathcal{W}[x]$ [38].

The limit in equation (3) can be estimated from the data set, as sketched in figures 1–3. To this end, we generate batches $X_i = \{x_1, \ldots, x_{d_{\text{batch}}}\}$ of size d_{batch}. Each x_j in X_i is extracted randomly between the minimum and maximum values of the trajectories Z_t. For each x_j, we consider all the n_j points Z_t^j, in all trajectories, which belong to the interval of width δ around x_j, see figure 3(a). The value of δ has to be chosen in such a way that all bins associated with the different x_j are sufficiently populated, ensuring the smoothness of the learned $\mu(x_j)$. We check *a posteriori* that the predicted dynamics, learned with such a δ, corresponds to the ground truth (see appendix A)[8].. We optimize μ_θ by minimizing the following loss function, cf figure 3(b)

$$L\left[\mu_\theta\right]\left(\theta\right) = \sum_{j=1}^{d_{\text{batch}}} \left| \mu_\theta\left(x_j\right) - \frac{1}{n_j} \sum_{Z_t^j} \Delta_1 Z_t^j \right|, \tag{4}$$

where $\Delta_1 Z_t \equiv (Z_{t+\mathrm{d}t} - Z_t)/\mathrm{d}t$. We consider the coarse-grained adimensional time t to correspond to the number of discrete-time updates of the system normalized by a suitable factor τ and thus $\mathrm{d}t = 1/\tau$.

In our data sets, the observed noise is often larger than the drift, cf figures 3(c) and 7(c), especially near the stationary state, where the drift coefficient vanishes altogether. This is why computing the targets $\frac{1}{n_j}\sum_{Z_t^j} \Delta_1 Z_t^j$ in equation (4) is essential. In fact, no learning would be possible without taking the targets to be arithmetic averages, due to the above-mentioned large fluctuations.

Since our task is to understand the order-parameter dynamics, we restrict ourselves to the problem of learning one-dimensional data. This allows for an efficient estimate of the drift coefficient in equation (2). In one dimension, the stochastic quantity Z_t indeed hits the different intervals sufficiently many times during the evolution, which is needed for proper sampling and computing $\mu(x)$. To reduce over-fitting, we train n_{train} different models $\mu_\theta^i(x)$, with the loss function (4). To each of these models, we assign a weight w_i equal to the inverse of the mean square error between the data estimate of $\mu(x)$ and the network result μ_θ^i. As a reference model μ_θ, we take the weighted average over this 'ensemble' of models:

$$\mu_\theta = \sum_{i=1}^{n_{\text{train}}} \frac{w_i\, \mu_\theta^i}{\sum_{i=j}^{n_{\text{train}}} w_j}. \tag{5}$$

Specifically, the values of $n_{\text{train}}, d_{\text{batch}}, n_{\text{epochs}}$ and δ we adopt for the considered models are reported in table A1 (see appendix A).

In order to learn the diffusion coefficient, we use the 'second moment' of $\mathrm{d}Z_t$, which is the quadratic variation $[Z]_t$. For stochastic processes as in equation (1), this is given by [39, 40]

$$\left[Z\right]_t = \int_0^t \mathrm{d}Z_s^2 = \int_0^t \mathrm{d}s \sigma^2\left(Z_s\right), \tag{6}$$

which is nothing but the integral version of the differential equation

$$\partial_t \left[Z\right]_t = \sigma^2\left(Z_t\right). \tag{7}$$

[8] This interval is defined as $B_j^\delta \equiv \{\text{All } Z_t\text{such that } |Z_t - x_j| < \delta\}$, and its cardinality is $n_j \equiv \#B_\delta^j$. We denote the points in this interval as Z_t^j.

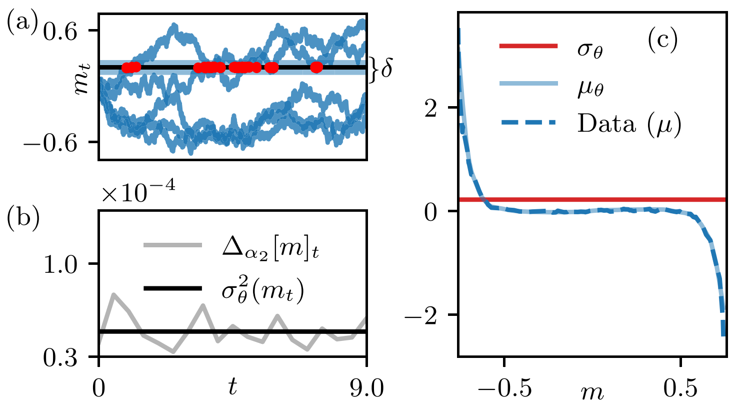

Figure 3. Estimation of the drift and diffusion coefficients μ and σ. In panel (a), some exemplary trajectories Z_t for the Ising model on a 128×128 square lattice are depicted. Here, the order parameter Z_t is the magnetization m_t. These trajectories are generated using the dynamical rules reported in equation (14) and provide the variable m_t. For a given m_t (black solid line), all the n_i intersections (red dots) within the shaded area are used to compute the quantity $\mu(x)\mathrm{d}t$. The latter is plotted in panel (c) (dashed line), while the solid line is obtained from the trained network $\mu_\theta(m)$. In panel (c) the network for the diffusion coefficient $\sigma_\theta(m)$ is also shown, in order to compare its dependence on the magnetization m with the one of the drift. This can also be seen in panel (b), where the values for the finite difference of the quadratic variation $\Delta_{\alpha_2}[m]_t$, upon which the network is trained, are provided together with the values of the corresponding predictions of the neural network $\sigma^2_\theta(m_t)$.

To train the network for the diffusion coefficient σ_θ, we devise a coarse-graining procedure that makes the spin-flip noise of the stochastic many-body dynamics look like a Wiener process. To this end, we first compute the quadratic variation from trajectories as

$$[Z]_t \cong \sum_{u\in[0,t]} \left(Z_{u+\alpha_1\mathrm{d}u} - Z_u\right)^2 . \tag{8}$$

Here, the integer factor $\alpha_1 \geqslant 1$ may allow one to magnify the variation at the different times. Furthermore, we approximate equation (7) by

$$\partial_t [Z]_t \approx \Delta_{\alpha_2} [Z]_t = \frac{[Z]_{t+\alpha_2\mathrm{d}t} - [Z]_t}{\alpha_2 \mathrm{d}t}. \tag{9}$$

The factor $\alpha_2 \gg 1$ allows one to coarse-grain the noise over many discrete time-steps, which proved necessary for convergence during the training procedure. This is mainly due to the fact that the finite difference in equation (9) is stochastic. For this reason we need an average in order to obtain valuable information for the training. Equation (9) will still be a good approximation of a time derivative if we consider a time window $\alpha_2 dt$ much smaller than the time during which relaxation to stationarity takes place. The optimization of the network parameters is then performed by minimizing the loss function

$$L[\sigma_\theta](\theta) = \sum_t |\Delta_{\alpha_2}[Z]_t - \sigma^2_\theta(Z_t)|. \tag{10}$$

Note that this loss function is insensitive to the sign of σ_θ. This is not a problem since the stochastic increment $\mathrm{d}W_t$ is symmetric under a change of sign.

To summarize, we explicitly selected the following hyper-parameters to control convergence: the number of hidden layers, the learning rate, the number of epochs, the batch size d_{batch}, and the number of models used to compute an 'average model' n_{train}. Additionally, we chose a threshold δ from which to compute finite increments from a given initial condition, and a time step $\mathrm{d}t$. We also set two constants α_1 and α_2. The constant α_1 is used in approximating the quadratic variation: $[Z]_t = \sum_u (Z_{u+\alpha_1\mathrm{d}u} - Z_u)^2$. The constant α_2 is used in computing its finite difference: $\sigma^2(Z_t) = ([Z]_{t+\alpha_2\mathrm{d}t} - [Z]_t)/\alpha_2\mathrm{d}t$. For further details on the training procedure, we refer to the appendix A and to table A1.

3. The kinetic ising model

3.1. The model and its dynamics

The Ising model is a paradigmatic model of statistical mechanics. It provides a qualitative description of the behavior of molecular magnetic dipoles in a metal. The crystalline structure of the metal is modeled as a lattice of N sites. At each site $i = 1, 2, \ldots, N$, a magnetic dipole is represented as a spin variable $s^i = \pm 1$. The spins interact with each other according to the following energy functional (Hamiltonian)

$$H(s) = -\frac{1}{2}\sum_{\langle ij\rangle} s^i s^j. \tag{11}$$

Here, the notation $\langle ij\rangle$ restricts the sites i and j in the sum to be nearest neighbors on the lattice. We consider a two-dimensional square lattice. This Hamiltonian presents a $\mathbf{Z}_2$ symmetry since it is invariant under sign change of every spin variable $s^i \to -s^i$. At thermal equilibrium at a given temperature T, each spin configuration has a probability described by the Boltzmann distribution $\pi_{\mathrm{B}}(s) \propto e^{-H(s)/k_{\mathrm{B}}T}$, where k_{B} stands for the Boltzmann constant. Given the magnetization

$$m = \frac{1}{N}\sum_i s^i, \tag{12}$$

the order parameter of the model is the expectation of the absolute value of m in the Boltzmann distribution. The system undergoes a continuous transition from an ordered phase with finite magnetization at sufficiently low temperatures, to a disordered one with vanishing magnetization. While in one dimension the model predicts a finite magnetization only at zero temperature, in two dimensions the critical temperature T_{c} corresponding to the phase transition is finite. Close to T_{c}, the value of the average magnetization $\bar{m}$ follows a power-law behavior

$$\bar{m} \propto |T - T_{\mathrm{c}}|^{\beta}, \tag{13}$$

where β is a so-called critical exponent.

The Ising model discussed above does not possess inherent dynamics. In order to apply our ML method to this model we can endow it with Glauber dynamics using Metropolis–Hastings sampling, which is usually utilized for sampling the Boltzmann distribution of the model. Such a dynamic is defined by the single spin-flip probabilities $P = P(s^i \to -s^i)$, updating the spin variables in the lattice according to

$$P\left(s^i_t \to -s^i_t\right) = \begin{cases} \exp\left(-\Delta E/k_{\mathrm{B}}T\right) & \text{if } \Delta E > 0 \\ 1 & \text{if } \Delta E \leqslant 0 \end{cases} \tag{14}$$

where $\Delta E = H(s^1_t, \ldots, -s^i_t, .., s^N_t) - H(s^1_t, \ldots, s^i_t, .., s^N_t)$ is the energy change associated with the transition. For the completion of a single discrete time step $s_t \to s_{t+1}$, a single spin-flip is attempted N times at a random site. For such a *dynamical* Ising model, the (stochastically evolving) order parameter m_t is defined as in equation (12) for an evolving configuration s_t. We choose each of the spins in the initial configuration to be up or down with equal probability, so that for large systems $m_0 \approx 0$. For further detail about the model and its field theoretical representation, see appendix B.1.

3.2. Neural network results

Given a set of trajectories for m_t at temperature T, we learn the corresponding drift term μ^T_θ using the approach explained above and the loss function $L[\mu^T](\theta)$ in equation (4). The drift term essentially acts as a directed force on the order parameter and it is thus natural to define an effective potential driving the motion of m_t via the integral

$$M^T_\theta(m) \equiv -\int_0^m \mathrm{d}x \mu^T_\theta(x)\,. \tag{15}$$

Our results reported in figure 4 show that upon increasing T the effective potential undergoes a transition from a functional form exhibiting a double well to a single well potential. This fact is connected with the equilibrium Ising phase transition which can be understood as follows. The stationary values of the expectation of the order parameter $\bar{m}_{\mathrm{stat}}$ correspond to the minima of the effective potential M^T_θ, see figure 4, and thus to zeroes of the drift coefficient, $\mu^T_\theta(\bar{m}_{\mathrm{stat}}) = 0$, cf figure 5. Since the considered discrete-time dynamics samples the Boltzmann distribution at stationarity, one essentially has that the stationary values $\bar{m}_{\mathrm{stat}}$ should approximate the equilibrium order parameter $\bar{m}$, thus connecting the retrieved potential to the Ising transition.

To benchmark the results from the trained networks μ^T_θ, we can thus extract the critical temperature T_{c} and the critical exponent β of the order parameter and compare them with the known values for the Ising

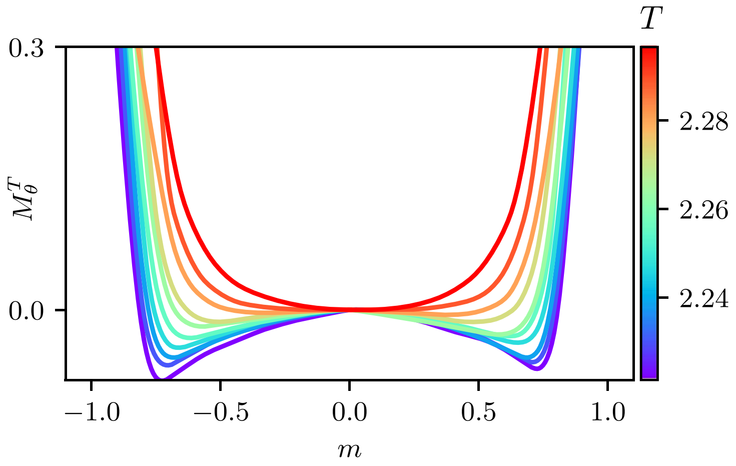

Figure 4. Effective potential M_θ^T of the Ising model. The effective potential M_θ^T in equation (15) is determined by the integral of the drift μ_θ with respect to the magnetization m. One can observe that for increasing temperature this effective potential changes its functional form from having two distinct minima in the double-well potential to having only one minimum at $m = 0$.

model. We fit the stationary magnetization $\bar{m}_{\mathrm{stat}}$ to the scaling form of equation (13) by minimizing the function

$$\epsilon\left(\tilde{c}_1, \tilde{T}_{\mathrm{c}}, \tilde{\beta}\right) = \sum_T |\mu_\theta^T\left(\tilde{c}_1|T-\tilde{T}_{\mathrm{c}}|^{\tilde{\beta}}\right)|^2. \tag{16}$$

The positive function $\epsilon(\tilde{c}_1, \tilde{\beta}, \tilde{T}_{\mathrm{c}})$ vanishes when $\tilde{c}_1|T-\tilde{T}_{\mathrm{c}}|^{\tilde{\beta}} = \bar{m}_{\mathrm{stat}}$. We consider the values of c_1, T_{c} and β that minimize ε in equation (16). To find them, the zeroes of the drift coefficient μ_θ^T, are computed using the exact derivatives via automatic differentiation. This is possible since we use differentiable neural networks. We find the following values: $\beta = 0.156 \pm 0.001$, $T_{\mathrm{c}} = 2.271 \pm 0.001$, and $c_1 = 1.076 \pm 0.002$. Note that the errors reported are only those related to the fit and do not consider finite-time and finite-size errors. For the Ising model, the analytical values are $\beta = 1/8$ and $T_{\mathrm{c}} = 2/\ln(1+\sqrt{2}) \cong 2.269$ [41]. Our results are thus in good agreement with the exact values and show that the networks are able to provide a sound description of the critical behavior encoded in the data they are trained on.

Close to the critical point, the Ising model with Glauber dynamics is expected to fall in the model A class according to the Halperin classification [42]. This is a pure relaxation model for a time dependent field in a double well potential, subject to uncorrelated white noise [43–46]. The latter feature is indeed reflected in our results on the learned diffusion coefficient σ_θ, shown in figures 3(b) and (c). There, we present σ_θ for $T = 2.269$, which is in proximity to the critical temperature. As can be seen, the diffusion coefficient σ_θ is essentially constant when compared with the drift coefficient, entailing white noise in the dynamics of m_t.

We thus showed how the learned networks are able to encode significant information about the statics, i.e. the order-disorder phase transition (see figure 5) and the dynamics, i.e. the form of the noise, for the process under investigation, through a simple equation.

4. The contact process

4.1. The model

We now apply our method to a paradigmatic nonequilibrium process, the so-called contact process [47, 48]. It was introduced to describe epidemic spreading in the absence of immunization. It is not defined via an energy function but solely via dynamical rules. The contact process shows a nonequilibrium continuous phase transition which belongs to the directed percolation universality class [49–53].

Within the epidemic spreading interpretation of the model, each lattice site i represents an individual which can either be found in the healthy state $s_t^i = 0$ (inactive site) or in the infected state $s_t^i = 1$ (active site). We consider here the case of a one-dimensional lattice. The dynamics occur in discrete time as follows: first, given the configuration s_t at time t, we calculate the probability that each spin flips through the rules

$$\begin{aligned} P\left[0 \to 1, n_t^i\right] &\propto \kappa \mathrm{d}t n_t^i/2\,, \\ P\left[1 \to 0, n_t^i\right] &\propto \gamma \mathrm{d}t\,. \end{aligned} \tag{17}$$

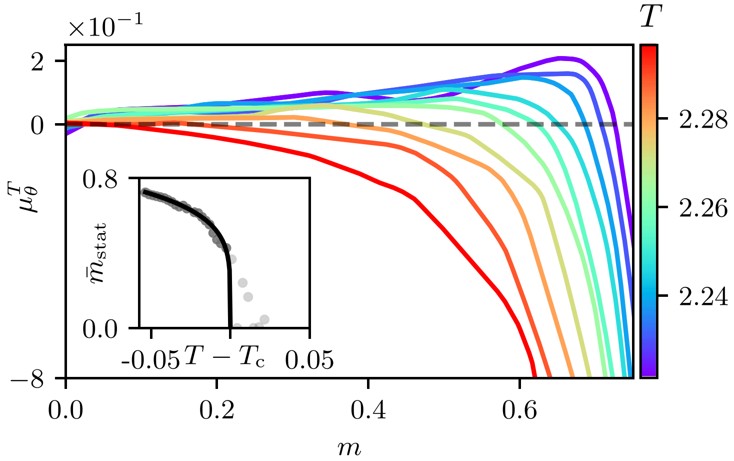

Figure 5. Neural networks for the drift coefficients μ_θ^T for the Ising model. Due to the $\mathbb{Z}_2$ symmetry of the Hamiltonian, only the behavior for $m > 0$ is shown. Warmer tones correspond to higher temperatures. The higher the temperature, the closer the non-trivial zero of μ_θ^T corresponding to $\bar{m}_{\rm stat}$ is to zero. In the main figure, the dashed line at zero intercepts the graph of the function μ_θ^T, at the stationary value $\bar{m}_{\rm stat}$ for each temperature T. Specifically, these values correspond to points where $\mu_\theta^T(\bar{m}_{\rm stat}) = 0$. In the inset, $\bar{m}_{\rm stat}$ values are shown as a function of their respective temperatures as grey bullets. Additionally, the inset displays the power law behavior of $\bar{m}_{\rm stat}$ relative to the distance from the critical temperature $T_{\rm c}$, derived by minimizing the function $\epsilon(\tilde{c}_1, \tilde{T}_{\rm c}, \tilde{\beta})$ in equation (16) as a solid black line.

Here, we introduced the healing rate γ, the infection rate κ and n_t^i indicates the number of infected nearest neighbors of i. Then, according to the above probability, a spin is extracted, and the corresponding flip is performed. The order parameter is the number density of infected sites

$$\rho_t = \frac{1}{N}\sum_i s_t^i, \tag{18}$$

with N being the total number of sites. We here consider $\rho_0 = 1$ as the initial value for the density (all sites infected).

From the dynamical rules in equation (17), one can see that the state with all healthy sites is a stationary state. In fact, this is a so-called absorbing state since it can be reached during the dynamics but it cannot be left. For any finite system, there is always a finite probability of hitting the absorbing state, which is the unique stationary state of the system. In the thermodynamic limit ($N \to \infty$) and for sufficiently large infection rates, a phase with a finite density of infected sites, usually called fluctuating phase [52–54], becomes stable. In finite systems, this phase eventually dies out and only appears within a meta-stable timescale. The absorbing phase and the fluctuating phase are separated by a continuous phase transition occurring at a finite critical value of the infection rate $\kappa_{\rm c}$, above which the system features a nonzero expectation of the stationary density $\bar{\rho}_{\rm stat}$. In proximity to the phase transition, the density follows a power-law behavior

$$\bar{\rho}_{\rm stat} \propto |\kappa - \kappa_{\rm c}|^\beta. \tag{19}$$

In the following, we focus on a one-dimensional lattice made of 100 sites and measure the infection rate in units of γ.

4.2. Neural network results

We start by discussing the results for the drift term of the contact process. As for the kinetic Ising model, we train the network for many data sets of trajectories. For each data set at infection rate κ, we train a model μ_θ^κ. The results for learned drifts μ_θ^κ are shown in figure 6, for different values of κ. Decreasing κ, the zero crossings $\mu_\theta^\kappa(\bar{\rho}_{\rm stat}) = 0$ occur at progressively smaller values of $\bar{\rho}_{\rm stat}$. In the inset, we illustrate how these can be used to extract the critical infection rate $\kappa_{\rm c}$ and the associated critical exponent β. As for the kinetic Ising model, we can fit the density of infected sites to the power law of equation (19) by minimizing the function $\epsilon(\tilde{c}_1, \tilde{\kappa}_{\rm c}, \tilde{\beta})$:

$$\epsilon\left(\tilde{c}_1, \tilde{\kappa}_{\rm c}, \tilde{\beta}\right) = \sum_\kappa \left|\mu_\theta^\kappa\left(\tilde{c}_1 |\kappa - \tilde{\kappa}_{\rm c}|^{\tilde{\beta}}\right)\right|^2. \tag{20}$$

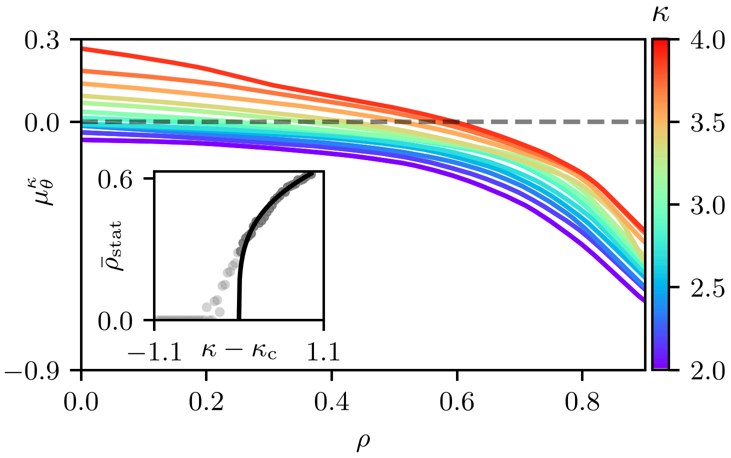

Figure 6. Drifts μ_θ^κ as a function of ρ, for different values of the infection rate. In the main figure, the drifts, which are color coded according to the infection rate κ, are depicted. The zeros of the drift term correspond to the stationary values of the stable stationary points $\bar{\rho}_{\rm stat}$, where $\mu_\theta^\kappa(\bar{\rho}_{\rm stat}) = 0$. The dashed line in the figure indicates where these zeros intersect with the drift coefficient. As the infection rate decreases, the stationary density collapses to $\bar{\rho}_{\rm stat} = 0$. From these values, represented as grey bullets in the inset, the critical exponent β can be extracted in the proximity of the critical infection rate $\kappa_{\rm c}$. The solid black line shown in the inset represents the power law behavior of $\bar{\rho}_{\rm stat}$ as a function of the infection rate κ. This power law is obtained by minimizing the function $\epsilon(\tilde{c}_1, \tilde{\kappa}_{\rm c}, \tilde{\beta})$ in equation (20). The drift terms in the active phase do not vanish when approaching $\rho = 0$ as one may expect. As discussed in the main text, this is a consequence of the fact that trajectories in the active phase do not visit sufficiently often values below the stationary values $\bar{\rho}_{\rm stat}$, which affects the training procedure.

The values $c_1, \kappa_{\rm c}$ and β that we find are $\beta = 0.28 \pm 0.03$, $\kappa_{\rm c} = 3.062 \pm 0.003$. These values should be compared with the values obtained by means of Monte Carlo or series expansion $\kappa_{\rm c} = 3.29785(8)$ [55–57], $\beta = 0.276486(8)$ [58, 59].

Albeit this agreement, there is in fact a problem with the shape of the learned drifts μ_θ^κ, as shown in figure 6. Given that the contact process features an absorbing state at density $\rho = 0$, one should expect that the drift vanishes for this density. This is evidently not the case here. The reason lies in the fact that the physics actually influences the way in which training data can be produced. In our case, we train the network considering trajectories starting from the state with all sites infected. For such initial condition and being in the active phase, the density of infected sites will decrease with time until it reaches a (meta)stable finite value around which it will fluctuate. This implies that during the learning process values of the density smaller than the (meta)stable one, including the absorbing-state value $\rho = 0$, are not visited sufficiently often. Therefore, it is not possible to appropriately learn the drift term below such values.

In figure 7, we report the results for the learned diffusion coefficient $\sigma_\theta(\rho)$, together with the network prediction for $\mu_\theta(\rho)$ and the time derivative of the quadratic variation $\Delta_{\alpha_2}[\rho]_t$, which the network learns (details on the network parameters are given in table A1). We consider a value for the infection rate, $\kappa = 3.36$, in the proximity of the critical point $\kappa_{\rm c}$. In contrast to the Ising model, where both phases above and below the critical point are fluctuating, the presence of an absorbing phase dictates that the diffusion coefficient must vanish at zero density. This means that the noise must be multiplicative. In fact, it can be proven that the diffusion coefficient is proportional to the square root of the density [54, 60, 61], which is a consequence of the central limit theorem and the fact that only active sites can contribute to fluctuations (for details we refer to appendix B.2). Both the learned diffusion coefficient σ_θ and the drift μ_θ are not constant and approach zero for small ρ, see figure 7(c). They are not strictly zero at $\rho = 0$ due to the above-discussed limitations of the learning procedure.

In figure 7(a) we show a selected trajectory, for which we display ρ_t, We see that σ_θ yields a time averaged value of the (coarse-grained) derivative of the quadratic variation $\Delta_{\alpha_2}[\rho]_t$ on which it was trained. Moreover, we also see that the learned noise vanishes as the system enters the absorbing state, i.e. $\rho_t = 0$ (cf figure 7(b)).

5. Conclusions

We have shown how to encode a simple stochastic equation in an artificial neural network and applied this method to two paradigmatic models of statistical mechanics, both in and out of equilibrium. Both studied systems, the kinetic Ising model and the contact process, exhibit a continuous phase transition which also is captured by the network. For both models we identified the critical point and retrieved the static critical exponent β.

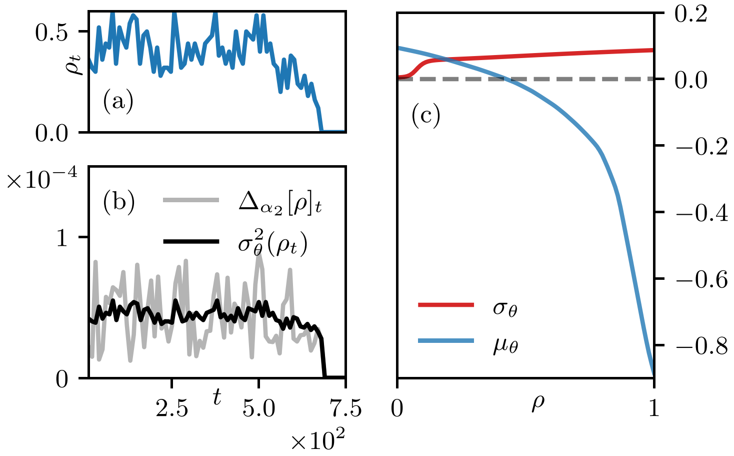

Figure 7. Diffusion in the contact process. (a) A sample trajectory ρ_t for $\kappa = 3.36$. Such trajectories ρ_t are used to compute the derivative of the quadratic variation $\Delta_{\alpha_2}[\rho]_t$. This time derivative is used to train the network σ_θ, via the loss in equation (10). (b) The network takes as input the values of ρ_t of a given trajectory and outputs the values $\sigma_\theta^2(\rho_t)$, which approximate the derivative of the quadratic variation $\Delta_{\alpha_2}[\rho]_t$, see equation (6). Unlike in the Ising model, σ_θ is not a constant function of the order parameter ρ_t. The reason is that in the absorbing phase (where the order parameter is strictly zero) no fluctuations take place, as shown the same trajectory in panel (a). (c) Learned diffusion coefficient σ_θ as a function of the density ρ. We observe that σ_θ indeed goes to zero for vanishing order parameter ρ, signalling multiplicative noise. For comparison, we also provide the results for the corresponding drift μ_θ.

It is important to note that within the chosen approach the network does not learn the order parameter from raw configurations. Rather, it is fed with a one-dimensional average value of an order parameter (density or magnetization) and outputs the one-dimensional drift and diffusion coefficients for a given order parameter value. The network thus learns one-dimensional quantities which simplifies the training process. In the case of the contact process, a multiplicative form of the noise is retrieved, while for the kinetic Ising model, the network learns a noise form that is approximately constant, i.e. independent of the value of the order parameter.

A natural future development would be to use the learned drift as a scaling function and to obtain all the critical exponents. This approach might also prove useful in classifying universal behavior of different processes, as two models are expected to belong to the same class, not only if they share the same set of critical exponents, but also if they share the same scaling function. Another point for future exploration is to go beyond the inherently Markovian assumption in equation (1), as the success of the results reported here, even under this assumption, could be attributed to the one-dimensional character of the training data. Future aims include the application of our approach to trajectories of open quantum processes and the utilization of machine learning methods that automatically infer the relevant order parameter [34].

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/fcarnazza/sde_order_params.

Acknowledgments

We thank M. Magoni for insightful discussions. We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-Number 2064/1-Project Number 390727645 (the Tübingen Machine Learning Cluster of Excellence), EXC Number 2075-Project Number 390740016 (the Stuttgart Cluster of Excellence SimTech), under Project No. 449905436, and through the Research Unit FOR 5413/1, Grant No. 465199066. This project has also received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101046968 (BRISQ). F. Carollo is indebted to the Baden–Württemberg Stiftung for the financial support of this research project by the Elite Programme for Postdocs. We acknowledge support from the Open Access Publication Fund of the University of Tübingen.

Table A1. Details of the networks' architecture and of the training procedure. All the trained networks are fully connected feed-forward multi-perceptron networks [63, 64]. In the first column, the number of neurons n_i in the i^{th} layer is reported, as $n_1 \times \ldots \times n_{\text{out}}$, with i in $\{1, 2, \ldots, \text{out}\}$. To train each network, the optimizer RMSprop algorithm in the PyTorch implementation is used, where only the learning rates were modified. The specific activation functions are also displayed. The quantity n_{train} refers to how many models are trained on each dataset, weighted averages of which are taken to compute the reference model. The dimension of the batch use to train is d_{batch}, while n_{epochs} is the number of epochs. The width of the interval from which μ is computed (cf figure 3) is δ, while the coarse graining of the discrete time is τ. The constant α_1 is used in computing the approximation for the quadratic variation $[Z]_t = \sum_u (Z_{u+\alpha_1 \mathrm{d}u} - Z_u)^2$. The constant α_2 is instead using in computing its finite difference: $\sigma^2(Z_t) = ([Z]_{t+\alpha_2 \mathrm{d}t} - [Z]_t)/\alpha_2 \mathrm{d}t$.

Network details						
Model	Layers architecture	Learning rate (RMSprop optimizer)	Activation function	n_{train}, d_{batch}, n_{epochs}	δ,τ	α_1, α_2
μ_θ Ising	$1 \times 50 \times 50 \times 1$	0.3×10^{-3}	ReLU	10, 100, 7000	0.01, 1000	—
σ_θ Ising	$1 \times 64 \times 1$	10^{-3}	Tanh (intra-layers), Sigmoid (output)	10, 100, 5000	—	1, 500
μ_θ Contact process	$1 \times 50 \times 1$	0.5×10^{-3}	ReLU	10, 100, 2000	0.05, 100	—
σ_θ Contact process	$1 \times 64 \times 1$	10^{-3}	Tanh (intra-layers), Sigmoid (output)	10, 100, 5000	—	10 100

Appendix A. Neural network details and integration of the learned stochastic equations

Because of the different properties of the two models considered in the present work, the kinetic Ising model and the contact process, the employed networks and the hyper-parameters adopted to train them are slightly different. In the following, we specify the details of the networks and how the integration of the Itô equation is performed. The code we use is available at [62].

A.1. Neural network and training details

We model the drift μ_θ as a fully connected feed-forward neural network. The network is trained by employing back propagation methods to optimize the loss function (4) This optimization minimizes the distance between the function $\mu_\theta(x)$ and the drift coefficient μ. The back propagation lets us compute the gradients used in an optimization routine. This routine requires as input a constant, namely, the learning rate, which amounts to the optimization step in the gradient descent algorithm. The order of magnitude of the learning rate should be small enough to learn the data's essential details yet not too small to avoid learning the noise effects. Moreover, lower learning rates make the optimization procedure slower. The learning rate we choose is thus a compromise between the optimization velocity and the accuracy of the results. We optimize the network to learn $\Delta_1 Z_t \tau$. The learned μ_θ then has to be multiplied with τ to make it comparable with the training data. For the Ising model, the time scale τ is set to 1000. For the contact process, it is 100. Similarly, the network σ_θ is a fully connected feed-forward network. As for μ_θ, the input and output dimensions are one-dimensional. Both for μ_θ and σ_θ the adopted optimizer is the PyTorch implementation of the RMSprop algorithm [65]. The architecture and training details for the networks and σ_θ are reported in the table A1, both for the kinetic Ising model and the contact process. For both of this processes the power law for the stationary values $\bar{m}_{\text{stat}}$ and $\bar{\rho}_{\text{stat}}$ only applies in the vicinity of the critical point, and only in the ordered and the active phase respectively. For the Ising model, the sum in equation (16) is computed for 15 values of the temperature equally spaced in between a minimum value $T_{\text{min}} = 2.2214$ and maximum value $T_{\text{max}} = 2.2759$. Similarly, for the contact process, in the sum in equation (20), we use 31 equally spaced values of the infection rate κ, from $\kappa_{\text{min}} = 2.0$ as lowest value to highest value $\kappa_{\text{max}} = 2.9831$.

To find the best critical values in equations (16) and (20), we use the minimization library [66] that allows to compute exact gradients on the (differentiable) networks μ_θ and σ_θ and minimize them numerically.

A.2. Integration of the Itô equation

In the present work, we extract an approximation to the drift coefficient μ and the time derivative of the quadratic variation $[Z]_t$ from the ground truth data. It is interesting to numerically solve the learned Itô equation (2) and compare the results with the ground truth Z_t. This can be readily done with the machine learning library Torchsde [67], which we adopt here. The numerical integration of Itô equations requires small time steps to achieve convergence. The fictitious time scale dt introduced to train the drift μ_θ thus comes in handy for the integration. One needs to pay attention to the increments $W_{t+u} - W_t$ of the Wiener

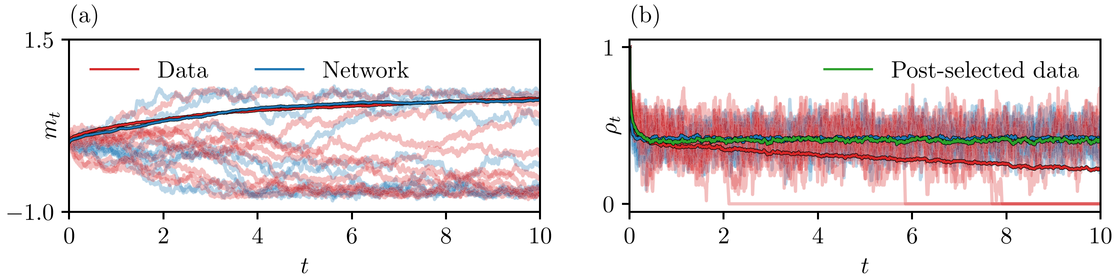

Figure A1. Dynamical trajectories for the kinetic Ising model (a) and the contact process (b). We report the ground truth trajectories ('Data', in red) as well as those generated by numerically integrating the learned Itô equation (2) ('Network', in blue). In panel (a), the time evolution of the magnetization of a two-dimensional lattice of size $N = 128 \times 128$ is shown for a temperature $T = 2.2214$. Some samples for the ground truth trajectories and the ones obtained from the network in lighter colors are displayed together with their averages over 100 trajectories in darker ones. Analogously, in panel (b), the time evolution of the density of active sites ρ_t on a one-dimensional chain of length $N = 100$ is shown for $\kappa = 3.36$, both for the ground truth and for the data generated by the network together with the respective order parameter estimated from 100 trajectories. The green line ('Post-selected data') represents the order parameter computed only from trajectories which do not decay into the absorbing state $\rho = 0$ (for the Ising model, the order parameter is the trajectory average of the absolute value of the magnetization m_t at time t).

process in equation (1), which should be distributed with a probability p following a normal distribution $\mathcal{N}(0,u)$ centered around zero and with variance u such that $p(W_{t+u} - W_t) = \mathcal{N}(0,u)$. The learned σ_θ, which is trained without using the time scale dt, has thus to be divided by $\sqrt{\mathrm{d}t}$ to make it comparable with the values obtained from the drift.

The results obtained by integrating the learned Itô equation (2) present a similar behavior to the ground truth dynamics, see figure A1(b).

For the contact process, our goal is to describe the dynamics up to the non-absorbing stationary state. For this reason, we restrict the training of the drift function μ_θ to $\rho_t > \delta$ in the data set, neglecting the region near the absorbing state ($\rho_t = 0$). This allows us to consider short trajectories while retaining important information about how the active (non-absorbing) stationary state ρ_{stat} is reached. This implies that the learned drift function μ_θ has only one stationary state, that is, a zero, in ρ_{stat}, but not in $\rho_t = 0$. When integrating the Itô equation, no trajectory thus goes to the absorbing state, something that instead happens to the ground truth data. The average of the ground truth data (referred to as 'Data' in figure A1) thus slowly decreases towards zero, while the average of only those trajectories in the ground truth data that do not end in the absorbing state exhibits a non-zero stationary state (indicated by 'Post-selected data' in figure A1). The latter agrees with the average obtained from integrating the learned Itô equation (indicated by 'Network' in figure A1). In this figure, the first moment of the dynamics is reported in darker tone both for the magnetization of the Ising model $\mathbf{E}[m_t]$ and the density of active sites in the contact process $\mathbf{E}[\rho_t]$. A possible measure $\mathbf{d}(m_t^{\text{data}}, m_t^{\text{net}})$ of the accuracy of the model is the averages difference between the first moment of the magnetization (density) m_t^{data} (ρ_t^{data}) and the one predicted by the neural network m_t^{net} over the time window $[0, T]$:

$$\mathbf{d}\left(m_t^{\text{data}}, m_t^{\text{net}}\right) = \frac{1}{T}\int_{[0,T]} \mathrm{d}t \left(\mathbf{E}\left[m_t^{\text{data}}\right] - \mathbf{E}\left[m_t^{\text{net}}\right]\right)^2. \tag{A.1}$$

In the reported figure, we find values $\mathbf{d}(m_t^{\text{data}}, m_t^{\text{net}} = 0.0013$ for the Ising model, and $\mathbf{d}(\rho_t^{\text{data}}, \rho_t^{\text{net}} = 0.0159$ in the case of the contact process.

Appendix B. Field-theoretic formulations

B.1. Kinetic Ising model

The field theory for the kinetic Ising model [42] is introduced by coarse-graining in space the originally discrete value of the spins s^i in the critical regime. Averaging over some mesoscopic spatial volume, a real-valued spin density field ψ is defined over a $(d+1)$-dimensional continuous space-time. Specifically, the configurations Σ_t have been coarse-grained so that at each point, a density $\psi \in \mathbf{R}$ is defined. Then the stochastic time evolution for the density ψ is provided by

$$\partial_t \psi(\mathbf{x}, t) = -\frac{\delta \mathcal{F}[\psi]}{\delta \psi(\mathbf{x}, t)} + \eta(\mathbf{x}, t), \tag{B.1}$$

with an effective potential functional of the form

$$\mathcal{F}[\psi] = \Gamma_0 \int \mathrm{d}^d \boldsymbol{x} |\nabla\psi(\boldsymbol{x},t)|^2 + u_0\psi^2(\boldsymbol{x},t) + r_0\psi^4(\boldsymbol{x},t)\,, \tag{B.2}$$

and a Gaussian noise with correlations

$$\langle \eta(\boldsymbol{x}_1,t_1)\,\eta(\boldsymbol{x}_2,t_2)\rangle = 2\Gamma_0\delta(\boldsymbol{x}_1-\boldsymbol{x}_2)\,\delta(t_1-t_2)\,. \tag{B.3}$$

The diffusion constant Γ_0 and the coupling constants in (B.2) are functions of the model parameters.

B.2. Contact process

For the field-theoretic formulation of the contact process, the average over some mesoscopic box in the lattice defines a coarse-grained density field $\rho(\mathbf{x},t)$ (instead of taking the average over the whole lattice $\bar{\rho}$). The Langevin equation for its time evolution can be derived directly from the master equation of the contact process and reads [54, 60]

$$\partial_t\rho(\mathbf{x},t) = D\nabla^2\rho(\mathbf{x},t) + \iota\rho(\mathbf{x},t) - \lambda\rho^2(\mathbf{x},t) + \zeta(\mathbf{x},t)\,. \tag{B.4}$$

The noise $\zeta(\mathbf{x},t)$ exhibits a multiplicative form

$$\langle \zeta(\boldsymbol{x}_1,t_1)\,\zeta(\boldsymbol{x}_2,t_2)\rangle = \Gamma\rho(\boldsymbol{x}_1,t_1)\,\delta(\boldsymbol{x}_1-\boldsymbol{x}_2)\,\delta(t_1-t_2)\,. \tag{B.5}$$

D is the diffusive constant, while the coupling constants ι, λ, and Γ are functions of the lattice details and of the infection rate. The occurrence of a term proportional to ρ and one proportional to ρ^2 in equation (B.4) can be explained heuristically from the mean-field treatment of the transition rates in equation (17). The number of sites becoming inactive at time t is $\sum_i s_t^i \propto \rho_t$. Instead, the number of sites becoming active is given by the number of inactive sites next to an active site that can be thus be infected. This number is given by $\sum_i (s_t^i - s_t^{i+1})^2 = 2\sum_i s_t^i - 2\sum_i s_t^i s_t^{i+1} \propto 2\rho_t - 2\rho_t^2$. In the mean-field treatment, the master equation thus reads $\partial_t\rho_t = (\kappa-1)\rho_t - \kappa\rho_t^2$. The form of the noise proportional to $\sqrt{\rho}$, can be justified by observing that only active sites contribute to the density fluctuations. To see this, let N be the total number of sites, and let n be the number of active sites at time t. The density of active sites at time t is thus $\rho_t = n/N$. Now, let X_i be the number of active sites at time $t' > t$ whose infection can be traced back to the ith active site at time t. Notice that the sequence $\{X_1,\dots,X_n\}$ is formed by independent identically distributed random variables, and $\rho_{t'} = 1/N\sum_i X_i$. Their expectation value and variance will thus be independent of the site i: $\mathbb{E}[X_i] = \nu$, $\mathrm{Var}[X_i] = \zeta^2$ for some real number ν and ζ. The relation between ν, ζ^2 and the sample average $\bar{X}_n = 1/n\sum_i X_i \equiv \rho_{t'}/\rho_t$ is described by the central limit theorem. This theorem states that for large n, the probability distribution p of the random variable $\sqrt{n}(\bar{X}_n - \nu)$ converges to a normal distribution centered around zero and with variance ζ^2, $\mathcal{N}(0,\zeta^2)$:

$$p\left(\sqrt{n}\,(\bar{X}_n-\nu)\right)\mathcal{N}\left(0,\zeta^2\right). \tag{B.6}$$

Note that for ρ_t to be finite also N must be large. Substituting $\sqrt{n} \to N\rho_t$ and $\bar{X}_n \to \rho_{t'}/\rho_t$, one obtains

$$p\left(\sqrt{N}(\rho_{t'}-\rho_t\nu)\right) \to \sqrt{\rho_t}\mathcal{N}\left(0,\zeta^2\right), \tag{B.7}$$

which means that the expectation value of $\rho_{t'}$ is $\rho_t\nu$, and its variance $\rho_t\zeta^2$.

ORCID iDs

Francesco Carnazza https://orcid.org/0000-0002-1458-8701
Federico Carollo https://orcid.org/0000-0002-6961-7143
Sabine Andergassen https://orcid.org/0000-0002-3128-6350
Georg Martius https://orcid.org/0000-0002-8963-7627
Miriam Klopotek https://orcid.org/0000-0001-9174-1282
Igor Lesanovsky https://orcid.org/0000-0001-9660-9467

References

[1] Van Kampen N 2007 *Stochastic Processes in Physics and Chemistry* 3rd edn (Elsevier)
[2] Coffey W T, Kalmykov Y P and Waldron J T 1996 *The Langevin equation* (World Scientific)
[3] Gillespie D 1992 *Markov Processes* (Academic)
[4] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087–92
[5] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864–71
[6] Rosenbluth M N and Rosenbluth A W 2004 *J. Chem. Phys.* **23** 356–9
[7] Widder C, Koch F and Schilling T 2022 *J. Chem. Phys.* **157** 194107
[8] Schilling T 2022 *Phys. Rep.* **972** 1–45
[9] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
[10] Mehta P, Bukov M, Wang C H, Day A G, Richardson C, Fisher C K and Schwab D J 2019 *Phys. Rep.* **810** 1–124
[11] Grogan F, Lei H, Li X and Baker N A 2020 *J. Comput. Phys.* **418** 109633
[12] Yang L, Gao T, Lu Y, Duan J and Liu T 2023 *Appl. Math. Model.* **115** 279–99
[13] Oh Y, Lim D and Kim S 2024 Stable neural stochastic differential equations in analyzing irregular time series data *12th Int. Conf. Learning Representations*
[14] Bai Y, Yan B, Zhou C, Su T and Jin X 2023 *Annu. Rev. Control* **56** 100909
[15] Glauber R J 1963 *J. Math. Phys.* **4** 294–307
[16] Süzen M 2014 *Phys. Rev.* E **90** 032141
[17] Walter J C and Barkema G 2015 *Physica* A **418** 78–87
[18] Majumdar S N 1999 *Curr. Sci.* **77** 370–5
[19] Callen H B 1985 *Thermodynamics and an Introduction to Thermostatistics* 2nd edn (Wiley)
[20] Halpin-Healy T 1998 *Phys. Rev.* E **58** R4096–9
[21] Halpin-Healy T 1991 *Phys. Rev.* A **44** R3415–8
[22] Xu J, Dupont E, Märtens K, Rainforth T and Teh Y W 2024 Deep stochastic processes via functional Markov transition operators *Proc. 37th Int. Conf. Neural Information Processing Systems* (*NIPS '23* Red Hook, NY, USA) (Curran Associates Inc.)
[23] Garnelo M, Rosenbaum D, Maddison C J, Ramalho T, Saxton D, Shanahan W, Teh Y W, Rezende D J and Eslami S M A 2018 Conditional neural processes (arXiv:1807.01613)
[24] Capobianco E 2008 *Physica* A **387** 4077–98
[25] Xu W, Chen R T Q, Li X and Duvenaud D 2022 Infinitely deep bayesian neural networks with stochastic differential equations *Proc. 25th Int. Conf. Artificial Intelligence and Statistics* (*Proc. Machine Learning Research*) vol 151, ed G Camps-Valls, F J R Ruiz and I Valera (PMLR) pp 721–38
[26] García C A, Félix P, Presedo J M R and Otero A 2022 *J. Comput. Phys.* **454** 110970
[27] Fernández-Fernández G, Manzo C, Lewenstein M, Dauphin A and Muñoz Gil G 2024 *Phys. Rev.* E **110** L012102
[28] Morrill J, Salvi C, Kidger P and Foster J 2021 Neural rough differential equations for long time series *Proc. 38th Int. Conf. Machine Learning, ICML (18–24 July 2021) Virtual Event* (*Proc. Machine Learning Research*) vol 139, ed M Meila and T Zhang (PMLR) pp 7829–38
[29] Kidger P, Morrill J, Foster J and Lyons T 2020 Neural controlled differential equations for irregular time series *Advances in Neural Information Processing Systems* vol 33, ed H Larochelle, M Ranzato, R Hadsell, M Balcan and H Lin (Curran Associates, Inc.) pp 6696–707
[30] Chen R T Q, Rubanova Y, Bettencourt J and Duvenaud D K 2018 Neural ordinary differential equations *Advances in Neural Information Processing Systems* vol 31, ed S Bengio, H Wallach, H Larochelle, K Grauman, N Cesa-Bianchi and R Garnett (Curran Associates, Inc.)
[31] Toth P, Rezende D J, Jaegle A, Racanière S, Botev A and Higgins I 2020 Hamiltonian generative networks *Int. Conf. Learning Representations*
[32] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Advances in Neural Information Processing Systems* vol 27, ed Z Ghahramani, M Welling, C Cortes, N Lawrence and K Weinberger (Curran Associates, Inc.)
[33] Kidger P, Foster J, Li X and Lyons T 2021 Efficient and accurate gradients for neural sdes *Advances in Neural Information Processing Systems*
[34] Dietrich F, Makeev A, Kevrekidis G, Evangelou N, Bertalan T, Reich S and Kevrekidis I G 2023 *Chaos* **33** 023121
[35] Clarke J B, Hastie J W, Kihlborg L H E, Metselaar R and Thackeray M M 1994 *Pure Appl. Chem.* **66** 577–94
[36] Pavliotis G A 2016 *Stochastic Processes and Applications* (Springer)
[37] Dynkin E B, Fabius J and Greenberg V 1965 Markov processes *(Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen, Band 121)* vol I (Springer)
[38] Øksendal B 2014 *Stochastic Differential Equations: an Introduction with Applications (Universitext)* 6th edn (Springer)
[39] Protter P 1990 *Stochastic Integration and Differential Equations* (Springer)
[40] Karandikar R L and Rao B V 2014 *Proc. Math. Sci.* **124** 457–69
[41] Onsager L 1944 *Phys. Rev.* **65** 117–49
[42] Hohenberg P C and Halperin B I 1977 *Rev. Mod. Phys.* **49** 435–79
[43] Debye P 1965 *Phys. Rev. Lett.* **14** 783–4
[44] Kawasaki K 1966 *Phys. Rev.* **145** 224–30
[45] Coniglio A and Zannetti M 1989 *Europhys. Lett.* **10** 575
[46] Majumdar S N, Bray A J, Cornell S J and Sire C 1996 *Phys. Rev. Lett.* **77** 3704–7
[47] Harris T E 1974 *Ann. Probab.* **2** 969–88
[48] Durrett R 1984 *Ann. Probab.* **12** 999–1040
[49] Broadbent S R and Hammersley J M 1957 *Math. Proc. Camb. Philos. Soc.* **53** 629–41
[50] Ódor G 2004 *Rev. Mod. Phys.* **76** 663–724
[51] Domany E and Kinzel W 1984 *Phys. Rev. Lett.* **53** 311–4
[52] Hinrichsen H, Jiménez-Dalmaroni A, Rozov Y and Domany E 1999 *Phys. Rev. Lett.* **83** 4999–5002
[53] Liggett T M 1985 *Interacting Particle Systems* (Springer)
[54] Hinrichsen H 2000 *Adv. Phys.* **49** 815–958
[55] Dickman R and Jensen I 1991 *Phys. Rev. Lett.* **67** 2391–4

[56] Dickman R and da Silva J K L 1998 *Phys. Rev.* E **58** 4266–70

[57] Jensen I and Dickman R 1993 *J. Stat. Phys.* **71** 89

[58] Muñoz M A, Grinstein G and Tu Y 1997 *Phys. Rev.* E **56** 5101–5

[59] Jensen I 1999 *J. Phys.* A **32** 5233–49

[60] Janssen H K 1981 *Z. Phys.* B **42** 151–4

[61] Cardy J L and Sugar R L 1980 *J. Phys.* A **13** L423

[62] 2023 (available at: https://github.com/fcarnazza/sde_order_params)

[63] Cybenko G 1989 *Math. Control Signals Syst.* **2** 303–14

[64] Rosenblatt F 1958 *Psychol. Rev.* **65** 386–408

[65] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Advances in Neural Information Processing Systems vol 32*, ed H ed Wallach, H Larochelle, A Beygelzimer, B F d'Alché, E Fox and R Garnett (Curran Associates, Inc.) pp 8024–35

[66] Feinman R 2021 Pytorch-minimize: a library for numerical optimization with autograd (available at: https://github.com/rfeinman/pytorch-minimize)

[67] Kidger P, Foster J, Li X and Lyons T J 2021 Neural sdes as infinite-dimensional gans *Proc. 38th Int. Conf. Machine Learning* (*Proc. Machine Learning Research*) vol 139, ed M Meila and T Zhang (PMLR) pp 5453–63