Dimension reduction of noisy interacting systems

Niccolò Zagli,[1,2,*] Grigorios A. Pavliotis,[3] Valerio Lucarini,[2,4] and Alexander Alecio[3]
[1]*Nordita, Stockholm University and KTH Royal Institute of Technology, Hannes Alfvéns väg 12, SE-106 91 Stockholm, Sweden*
[2]*Centre for the Mathematics of Planet Earth, University of Reading, Reading RG6 6AX, United Kingdom*
[3]*Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom*
[4]*Department of Mathematics and Statistics, University of Reading, Reading RG6 6AX, United Kingdom*

(Received 16 July 2022; accepted 4 January 2023; published 3 February 2023)

We consider a class of models describing an ensemble of identical interacting agents subject to multiplicative noise. In the thermodynamic limit, these systems exhibit continuous and discontinuous phase transitions in a, generally, nonequilibrium setting. We provide a systematic dimension reduction methodology for constructing low-dimensional, reduced-order dynamics based on the cumulants of the probability distribution of the infinite system. We show that the low-dimensional dynamics returns the correct *diagnostic* properties since it produces a quantitatively accurate representation of the stationary phase diagram of the system that we compare with exact analytical results and numerical simulations. Moreover, we prove that the reduced order dynamics yields also the *prognostic*, i.e., time-dependent properties, as it provides the correct response of the system to external perturbations. On the one hand, this validates the use of our complexity reduction methodology since it retains information not only of the invariant measure of the system but also of the transition probabilities and time-dependent correlation properties of the stochastic dynamics. On the other hand, the breakdown of linear response properties is a key signature of the occurence of a phase transition. We show that the reduced response operators capture the correct diverging resonant behavior by quantitatively assessing the singular nature of the susceptibility of the system and the appearance of a pole for real values of frequencies. Hence, this methodology can be interpreted as a low-dimensional, reduced order approach to the investigation and detection of critical phenomena in high-dimensional interacting systems in settings where order parameters are not known.

DOI: 10.1103/PhysRevResearch.5.013078

I. INTRODUCTION

The investigation of dynamical phenomena in complex networks constructed according to different topologies is an extremely active research area [1–3]. Interacting agent based models are commonly employed to model various phenomena in the natural sciences, social sciences and engineering [4,5], such as cooperation [6], synchronization [7], systemic risk [8], and consensus formation [9]. Several algorithms for sampling, optimization and the training of neural networks can be interpreted as interacting particle systems [10–12]. In the thermodynamic limit, such models often exhibit phase transitions as a result of the complex interplay between the interacting dynamics and the noise. Clearly, singularities associated to phase transitions, such as the divergence of correlation properties [6] and the breakdown of linear response properties [13,14], can only be observed in the mean-field (thermodynamic) limit. Consequently, their investigation involves the study of a (nonlinear and nonlocal) mean-field Fokker-Planck equation or a *brute force* approach, i.e., extensive numerical simulations of very large ensemble of agents. Reduction of complexity can be achieved by defining collective variables (reaction coordinates) able to accurately describe the full dynamics in a low-dimensional space. Nonetheless, while order parameters like magnetization can in many cases easily deduced for equilibrium systems using, e.g., symmetry arguments, the definition of reaction coordinates for nonequilibrium system is far more challenging [15–17].

The goal of this paper is to present a *model reduction* approach for the study of such infinite systems based on a systematic approximation of the full infinite-dimensional dynamics in terms of a low number of ODEs. This methodology can be applied to any interacting systems model with mean-field polynomial dynamics, with numerous applications including cooperation phenomena [6], synchronization of nonlinear, possibly chaotic, oscillators [18,19] and emergent phenomena in neural networks and life sciences [20,21]. The dimension reduction procedure we propose is based on a suitable closure method of the infinite hierarchy of equations for the moments or, equivalently, cumulants of the probability distribution of the infinite-dimensional system. Such closure method results in a deterministic parametrization of the full dynamics in terms of a low number of cumulants. One could potentially improve on this by using the Mori-Zwanzig formalism [22,23] to construct a stochastic, possibly non-Markovian, parametrization [24–26]. From a data-driven

*niccolo.zagli@su.se

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

perspective, one could rely on empirical model reduction [27] techniques to obtain closures from partial observations of the system. The resulting closure structure is given in terms of multilayer stochastic systems whose relevance and robustness has also been highlighted from an alternative, theory-informed parametrization perspective [28]. As validation case studies, we apply our dimension reduction methodology to investigate the nonequilibrium continuous phase transition in a model featuring noise-induced stabilization phenomena [29] and a model featuring an equilibrium discontinuous transition [30].

II. THE CLASS OF MODELS

We consider a system of exchangeable weakly interacting one-dimensional diffusions whose dynamics is governed by the following Stratonovich SDE:

$$dx_i = \left[F_\alpha(x_i) - \frac{\theta}{N}\sum_j^N \mathcal{U}'(x_i - x_j)\right]dt + \sigma(x_i)\circ dW_i \quad (1)$$

with initial condition $x_i \sim \hat{\rho}(x)$ and $i = 1, \ldots, N$. Each agent undergoes an internal dynamics given by the vector field $F_\alpha(x)$, depending on a set of parameters α, and is coupled with all the other agents through a symmetric interaction potential $\mathcal{U}(x) = \mathcal{U}(-x)$, with θ denoting the interaction strength. Furthermore, dW_i, $i = 1, \ldots, N$, are independent Brownian motions and $\sigma(x) > 0$ $\forall x \in \mathbb{R}$ is a multiplicative diffusion coefficient. The main assumption in this paper is that $F(x)$, $\mathcal{U}(x)$ and the diffusion matrix $\Sigma(x) = \sigma^2(x)$ all have a polynomial functional form. We consider quadratic interactions, $\mathcal{U}(x) = \frac{x^2}{2}$, corresponding to cooperative interactions among the agents that attempt to synchronise them towards their common center of mass $\bar{x}(t) = \frac{1}{N}\sum_i^N x_i(t)$. We are interested in the thermodynamic limit $N \to +\infty$ of Eq. (1). It is known that the empirical measure $\rho_N = \frac{1}{N}\sum_i^N \delta_{x_i(t)}$ converges (weakly) [31–33] to the one particle distribution $\rho(x,t)$ satisfying the nonlinear, nonlocal Fokker-Planck (McKean-Vlasov) PDE that, according to our setting, can be written as

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left(\frac{\sigma^2(x)}{2}\rho\frac{\partial}{\partial x}(f_{\langle x\rangle}(x) + \ln\rho)\right), \quad (2)$$

where $\rho(x,0) = \hat{\rho}(x)$ and

$$f_{\langle x\rangle}(x) = 2\int^x \frac{-\hat{F}_\alpha(y) + \theta(y - \langle x\rangle)}{\sigma^2(y)}dy + \ln\sigma^2(x). \quad (3)$$

$\langle x\rangle = \int_{\mathbb{R}} y\rho(y,t)dy$ represents the first moment of the distribution $\rho(x,t)$ and $\hat{F}_\alpha(x) = F_\alpha(x) + \frac{1}{2}\sigma(x)\sigma'(x)$. Equation (2) exhibits, at low temperatures, nonuniqueness of stationary solutions, that correspond to phase transitions [30,34].

Stationary solutions of Eq. (2) can be written as a one parameter family of distributions

$$\rho_0(x;m) = \frac{e^{-f_m(x)}}{\int_{\mathbb{R}} e^{-f_m(x)}dx} \equiv \frac{e^{-f_m(x)}}{Z(m)} \quad (4)$$

where the parameter m satisfies the self-consistency equation

$$m = R(m) \equiv \int_{\mathbb{R}} x\rho_0(x;m)dx \quad (5)$$

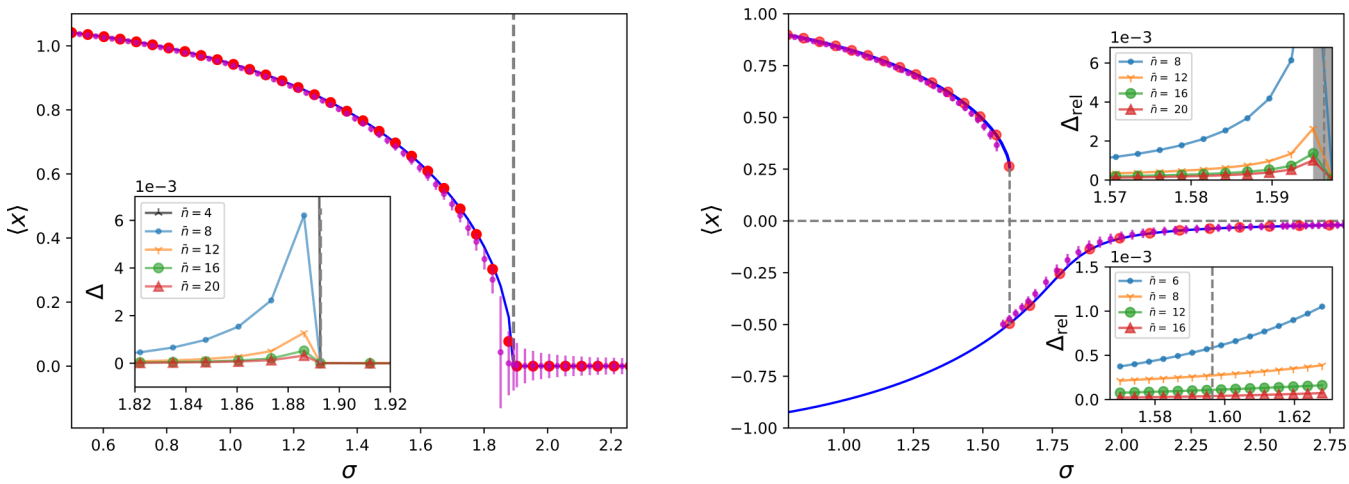
and $Z(m) > 0$ denotes the partition function. Equation (5) plays a major role in determining the stationary properties of the system. Solutions $m^\star$ of Eq. (5) correspond to stationary measures $\rho_0(x;m^\star)$ with first moment $\langle x\rangle = m^\star$, a suitable order parameter of the system for this type of quadratic interactions. Partial information on the stability of the invariant measures can be obtained by the investigation of the slope of the self-consistency equation $R'(m^\star) = \frac{dR(m)}{dm}|_{m^\star}$. In particular, if $R'(m^\star) > 1$, the stationary solution $\rho(x;m^\star)$ is unstable.

III. REDUCED ORDER DYNAMICS

In order to construct the reduced order dynamics, we multiply Eq. (2) by x^n, $n \in \mathbb{N}$, and integrate over $\mathbb{R}$. Given our assumptions on the drift and diffusion terms, this procedure results in an infinite hierarchy of equations for the moments $M_n = \langle x^n\rangle$ of ρ, see Appendix B for more details. In order to elucidate the procedure above, we will first consider model A defined by $F_\alpha(x) = -V'_\alpha(x)$ where $V_\alpha(x) = \frac{x^4}{4} - \alpha\frac{x^2}{2}$ is a double well potential if $\alpha > 0$ and the diffusion matrix is $\Sigma(x) = \sigma^2 + \sigma_m^2 x^2$. This model was introduced in [29] to investigate the effect of multiplicative noise on spatially extended systems. We mention that, if $\sigma_m = 0$, model A becomes the well-known Desai-Zwanzig model [35], a paradigmatic example featuring an equilibrium continuous phase transition. The state-dependent noise arises as the parameter α is not known with infinite precision and is allowed to randomly fluctuate in time, namely $\alpha \to \alpha + \sigma_m d\xi$ where $d\xi$ is another uncorrelated Brownian motion. Model A shows a noise induced stabilization phenomenon. When $\sigma_m \neq 0$, the multiplicative noise has a rectifying effect, pushing, for strong enough coupling θ, the phase transition point to higher and higher σ, see Appendix A and in particular Fig. 3 for more details. We apply the procedure mentioned at the beginning of this section to model A and obtain the following equations for the moments M_n

$$\begin{aligned}\frac{dM_n}{dt} =& n\left(\alpha - \theta + \frac{n}{2}\sigma_m^2\right)M_n - nM_{n+2} \\ &+ \frac{n(n-1)}{2}\sigma^2 M_{n-2} + n\theta M_1 M_{n-1}\end{aligned} \quad (6)$$

with $M_0 = 1$, $M_{-1} \equiv 0$. Firstly, we observe that the global coupling among the agents gives rise to an interaction term between the order parameter $\langle x\rangle = M_1$ and all the other moments M_n, introducing a nonlinear term in the hierarchy for the moments. Secondly, the nonlinear features of the dynamics given by $V_\alpha(x)$ introduce a (linear) dependence of lower moments on higher degree ones. The infinite hierarchy of moment Eq. (6) is equivalent to Eq. (2) and no reduction in the level of complexity of the mathematical description has been accomplished yet. The necessity of finding appropriate closure schemes for the hierarchy arises. Were we to truncate the system of Eq. (6) at a specific level $\bar{n}$, a closure scheme for $M_{\bar{n}+1}$, $M_{\bar{n}+2}$ in terms of M_n with $n < \bar{n}$ is needed. Truncated moment problems and closure schemes are not easily amenable to a mathematical investigation and are known to introduce statistical assumptions whose validity is difficult to justify, if not from an a posteriori perspective, see Refs. [36–38].



(a) (b)

FIG. 1. Phase diagram, $\langle x\rangle = \langle x\rangle(\sigma)$. The continuous blue line refers to the self-consistency equation, the red dots to the reduced order dynamics ($\bar{n} = 4$) and the magenta dots (with errorbars) to the numerical integration. (a) Continuous transition given by model A. The inset at the bottom shows the absolute error Δ between the reduced order dynamics and the self-consistency approach. The error Δ for $\bar{n} = 4$ is out of scale and peaks at a value $\Delta \approx 0.1$. The vertical dashed line refers to the critical condition $R'(0) = 1$. Fixed parameters are $(\alpha, \theta, \sigma_m) = (1, 4, 0.8)$. (b) Discontinuous phase transition of model B. The insets show the relative error $\Delta_{\rm rel}$ between the reduced order dynamics and the self-consistency approach. The inset at the top (bottom) refers to the upper (lower) branch of the phase diagram. The vertical dashed line is obtained numerically through the self-consistency approach and its value has been consistently checked to yield a slope $R'(m)$ such that $R'(m) - 1 \approx 10^{-4}$. Fixed parameters are $(\alpha, \theta, \mu) = (1, 4, 0.02)$.

Following [35,39,40], we implement a cumulant truncation scheme [40–42]. We introduce the cumulants k_n as

$$\sum_{n=1}^{\infty} k_n(t)\frac{\lambda^n}{n!} = \ln \int_{\mathbb{R}} \rho(x,t)e^{\lambda x}dx. \tag{7}$$

The truncation scheme consists of imposing the condition $k_{\bar{n}+1} = k_{\bar{n}+2} = 0$. This procedure provides a closure relations for $\bar{M}_{\bar{n}+1} = \bar{M}_{\bar{n}+1}(M_1, \ldots, M_{\bar{n}})$ and $\bar{M}_{\bar{n}+2} = \bar{M}_{\bar{n}+2}(M_1, \ldots, M_{\bar{n}})$. Alternatively, one can obtain from (7) and (2) an infinite hierarchy of equations for the cumulants

$$\frac{dk_n}{dt} = G_n(k_1, \ldots, k_n, k_{n+1}, k_{n+2}), \tag{8}$$

where the explicit expression of the nonlinear function $G_n(\cdot)$ is written Appendix B. Equation (8) indicates that the cumulant truncation scheme corresponds to a parametrization of the dynamics given by Eq. (1), in the limit $N \to +\infty$, in terms of a finite number $\bar{n}$ of cumulants. It is well known that such a scheme is inconsistent, since a function with a finite cumulant expansion cannot be positive if the order of the highest cumulant is larger than two [43]. Heuristically, a parametrization in terms of cumulants is expected to perform better than parametrizations in terms of (central) moments based on the observation that a Gaussian distribution has vanishing cumulants $k_n = 0$ for $n > 2$, while all (central) moments are nonzero. For non-Gaussian distributions, one expects that neglected higher-order cumulants will be smaller than the corresponding (central) moments. Moreover, the relevance of cumulants in the description of statistical properties of complex systems, especially in settings with athermal noise, has recently been highlighted, see Refs. [44,45] and references therein. We refer the reader to Appendix C for the comparison between different parametrizations and the validation of the cumulant truncation scheme for the systems under investigation. For model A, Eq. (5) predicts that the stable solution $\langle x\rangle = 0$ bifurcates when $R'(0) = 1$ through a continuous phase transition in two symmetric, competing states with opposite order parameter. Panel (a) of Fig. 1 shows the continuous phase diagram for the state with positive order parameter, obtained with the exact self-consistency equation and the reduced order dynamics, see Eq. (8). As soon as $\bar{n} = 4$ cumulants (main panel) are introduced, the reduced dynamics provides a very good approximation of the phase diagram. The critical value of the parameter is underestimated by the reduced order dynamics, with such approximation getting progressively better as more cumulants are considered. The accuracy of the reduced dynamics has been quantitatively assessed in terms of the absolute error Δ (shown in the inset) with respect to the self-consistency approach. The reduced dynamics has also been compared to numerical simulations of an ensemble of $N = 12\,000$ agents described by Eq. (1). We have used the Milstein scheme [46], which has strong order of convergence 1, with time step $\Delta t = 0.01$ and estimated the order parameter as the time average, at stationarity, of the center of mass $\bar{x}(t)$. Moreover, the reduced order dynamics has been initialised with a Gaussian initial condition, such that $(k_1, k_2) = (0.1, 0.01)$ and all others cumulants set to zero. Very good agreement is observed. Close to the phase transition, finite size effects arise in the numerical simulations. Noise-induced transitions among the two symmetric solutions become a relevant feature and one should consider the *rectified* order parameter (shown in the figure), obtained as the time average of $\bar{x}(t)$ conditioned on the fact that the system is in the basin of attraction of the positive solution. We have also probed the validity of the cumulant

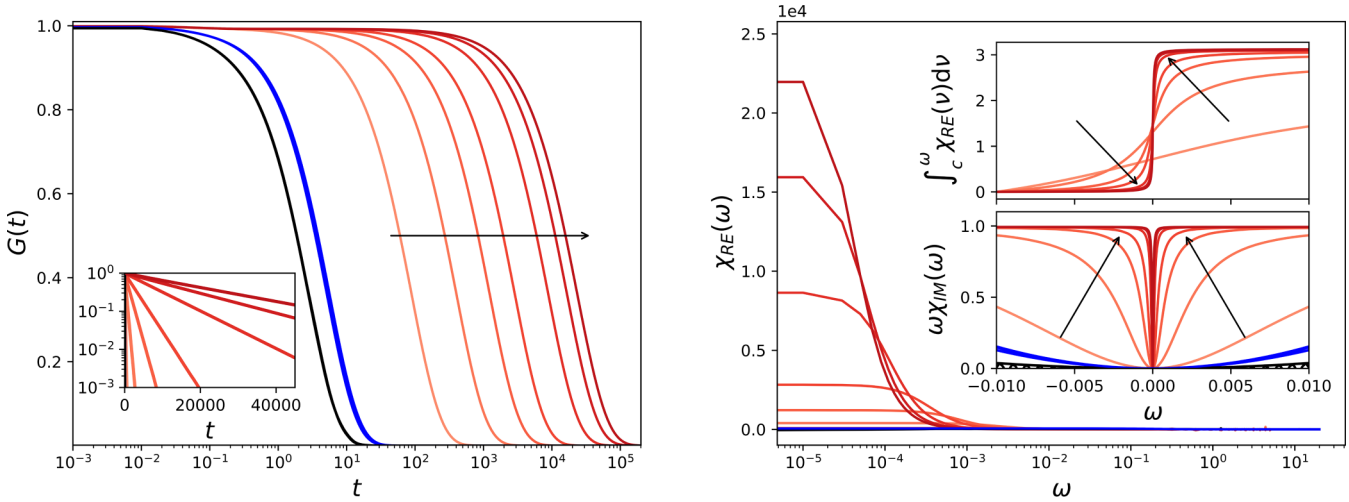


FIG. 2. Green function $G(t)$ (a) and susceptibility $\chi(\omega)$ (b) for model A. The blue (black) lines refer to a noncritical setting 5% below (above) the transition point. Red lines refer to critical settings. Fixed parameters are as in (a) of Fig. 1. The red color code and the arrows correspond to increasing values of $\bar{n} = 4, 6, 8, 10, 14, 18, 22$. In (b), black and blue lines have been multiplied by a scaling factor for graphical purposes.

based parametrization by investigating discontinuous phase transitions. We introduce model B, characterised by a tilted potential $V_{\alpha,\mu} = V_\alpha(x) + \mu x$ and additive noise $\Sigma(x) = \sigma^2$. Stationary properties of the reduced dynamics, with Gaussian initial condition $(k_1, k_2) = (1, 0.01)$, are in very good agreement with the other two approaches, see Panel (b). The insets show the relative error $\Delta_{\rm rel}$ between the reduced dynamics and the self-consistency equation. The top one, referring to the top branch of the phase diagram, shows that in the very close proximity, represented as a shaded area, of the transition point, $\Delta_{\rm rel}$ jumps to higher values, due to the fact that the reduced dynamics underestimates the critical value of the parameter and approaches it from below as $\bar{n}$ increases. The bottom inset shows that $\Delta_{\rm rel}$ for the bottom branch of the phase diagram is instead a smooth function that is not affected by the transition. This confirms that the reduced dynamics is able to track, as σ is parametrically changed, the disappearing attractor until a transition occurs to the other stable, smoothly changing, attractor. Noise-induced transitions are observed close to the phase transition in the finite system. Due to the asymmetry between the two competing states, the metastable lifetime of the state with $\langle x \rangle > 0$ decreases as the transition is approached and the system, after a short time, is driven to the other state of much longer lifetime.

The above results confirm that the reduced order dynamics correctly retains information of the exact invariant measure of the system. Below, we show that the approximate dynamics also captures time-dependent properties, and, specifically, correlations, by investigating, in the spirit of the fluctuations dissipation theorem [47,48], to time-modulated external perturbations. We report linear response properties of the reduced dynamics for model A (see Appendix A for response properties of model B). We perturb a stable stationary state with a homogeneous perturbation in the drift term $F(x) \to F(x) + \varepsilon T(t)$, where ε is small. Such procedure results in a one-cumulant perturbation $k_1^{(0)} \to k_1^{(0)} + \varepsilon T(t)$ for Eq. (8), where $k_1^{(0)}$ is the unperturbed order parameter. Following [14], we choose as temporal modulation for the forcing a Dirac's δ: $T(t) = \delta(t)$, which corresponds to a broad band forcing in frequency space. We then observe the Green function $G(t)$, associated to the order parameter, defined by $k_1(t) = k_1^{(0)} + \varepsilon \int_0^\infty G(t-\tau)T(\tau)d\tau$. Convergence to the linear regime has been assessed evaluating the response for different values of ε. Panel (a) of Fig. 2 shows that, at the transition point (red lines), the Green function has an exponential decay (see inset) with an associated timescale that is order of magnitudes greater than what is observed in noncritical settings (blue and black lines). Moreover, such timescale is an increasing function of the level of truncation $\bar{n}$ of the reduced dynamics, whereas no dependence on $\bar{n}$ is observed for the noncritical Green functions—see panel (a) of Fig. 2. The critical behavior is linked to the breakdown of linear response theory at the phase transition point, in the thermodynamic limit of Eq. (1) due to the agent-to-agent interactions, thus being associated with endogenous dynamical processes [13]. As the number of agents N is increased, one observes an emerging singular behavior in the susceptibility $\chi(\omega)$, defined as the Fourier transform of $G(t)$, signalled by a development of a pole ω_0 on the real axis of the frequencies [14]. The infinite hierarchy (7) corresponds to the thermodynamic limit of the ensemble of agents and one expects a diverging response in critical settings. However, we observe that the truncation scheme introduces a mollifying effect of the singular behavior of the reduced response operators. The resonance of such operators can be investigated through the susceptibility of the reduced dynamics that can be written as $\chi(\omega) = \frac{\kappa}{\omega - \omega_0 + i\gamma(\bar{n})} + r(\omega)$ where $\omega_0 = 0$ and $r(\omega)$ is an analytic function in the upper complex ω plane. As the number of reaction coordinates increases, $\bar{n} \to \infty$, the regularising effect vanishes, $\gamma(\bar{n}) \to 0$, and the susceptibility develops a singular behavior given by $\lim_{\bar{n}\to\infty} \chi(\omega) = -i\pi\kappa\delta(\omega - \omega_0) + \kappa\mathcal{P}(\frac{1}{\omega-\omega_0}) + r(\omega)$. Panel (b) confirms the appearance of an emerging pole with an

imaginary residue $\kappa = i|\kappa|$. The real part $\chi_{\rm RE}$ (main panel) of the susceptibility clearly shows the resonant δ-like behavior for $\omega = \omega_0$. Alternatively, the top inset shows that the primitive function of $\chi_{\rm RE}$ close to the pole ($c = -0.01$) converges accordingly to a Heaviside function. We observe that $\bar{n} = 4$ does not show a resonant behavior, even though it is associated with a longer timescale. The imaginary part $\chi_{\rm IM}(\omega)$ of the susceptibility (bottom inset), behaving like a Cauchy principal value distribution, yields a quantitative estimate $|\kappa| \approx 1$ for the residue of the pole. It is possible to obtain a formula for the amplitude of the residue as $|\kappa| = \frac{1}{\theta\tau_{x,A}}$, where (see Appendix D)

$$\tau_{x,A} = \frac{\int_0^{+\infty} \left\langle x(t)\arctan\left(\frac{\sigma_m}{\sigma}x(0)\right)\right\rangle_0 dt}{\left\langle x\arctan\left(\frac{\sigma_m}{\sigma}x\right)\right\rangle_0}. \tag{9}$$

Numerical simulations on an ensemble of $N = 16\,000$ agents yield a value of $\tau_{x,A} \approx 0.25$ and, since $\theta = 4$, $|\kappa| \approx 0.99$, validating thus our results. We remark that the existence of the pole ω_0 at the phase transition, as opposed to its residue κ, depends neither on the forcing $T(t)$ nor on the choice of the observable and can be related to spectral properties of suitably defined evolution operators [14]. This crucial property validates the use of our cumulant based reduced dynamics to settings where the order parameter is not known or cannot easily be written in terms of the cumulants.

IV. CONCLUSIONS

In this paper, we considered a class of models describing an ensemble of N identical interacting agents subject to multiplicative noise that exhibits phase transitions in the thermodynamic limit. We derived a reduced low-dimensional system for the moments of the probability distribution function of the mean-field dynamics. We showed that such approximate dynamics provides an accurate representation of the stationary phase diagram, even for a very low number (e.g., 4) of moments. This indicates that the cumulants act as effective reaction coordinates, which are able to capture the essential properties of the system with moderate loss of information due to the cumulant truncation. Additionally, the linear response properties of the projected dynamics agrees with that of the full system, and the breakdown of the corresponding linear response operators can be used to characterise the phase transition occurring in the system. Hence, our methodology seems useful for performing linear stability analysis for a large class of interacting multiagent systems, and for predicting their response to forcings of general nature. It is worth investigating how our dimension-reduction methodology compares with what one would obtain by applying variational autoencoders [49] to construct a surrogate, low-dimensional representation of the system. On top of the detection of critical phenomena for high-dimensional systems, a further application of our methodology relates to the issue of parameter estimation for interacting systems. Current parameter estimation techniques rely on suitable fitting procedures of the observational data to the infinite-dimensional dynamics [50], whereas one could envision simpler settings where the reduced order dynamics is taken as the reference point. We expect that this complex reduction methodology will not prove to be as effective when the system does not exhibit a clear separation of time or phase space scales, see Ref. [51] and references therein for a review of systems that can be "effectively reduced" either from a theoretical or data-driven perspective.

ACKNOWLEDGMENTS

VL acknowledges the support received by the European Union's Horizon 2020 research and innovation program through the project TiPES (Grant Agreement No. 820970). The work of GP was partially funded by the EPSRC, Grant No. EP/P031587/1, and by J.P. Morgan Chase & Co through a Faculty Research Award No. 2019 and 2021. NZ has been supported by an EPSRC studentship as part of the Centre for Doctoral Training in Mathematics of Planet Earth (Grant No. EP/L016613/1) and by the Wallenberg Initiative on Networks and Quantum Information (WINQ).

APPENDIX A: THE MODELS

In this section, we provide further details on the models we have studied in the paper. As specified in the main text, we investigate multi agent systems whose dynamics is given by the following equations:

$$dx_i = \left[F_\alpha(x_i) - \frac{\theta}{N}\sum_j^N \mathcal{U}'(x_i - x_j)\right]dt + \sigma(x_i)dW_i, \tag{A1}$$

where $i = 1, \ldots, N$. The examples we have provided refer to a quadratic interaction potential $\mathcal{U}(x) = \frac{x^2}{2}$. This results in

$$dx_i = [F_\alpha(x_i) - \theta(x_i - \bar{x})]dt + \sigma(x_i)dW_i, \tag{A2}$$

where $\bar{x}(t) = \frac{1}{N}\sum_i^N x_i(t)$ is the common center of mass of the system. Given that the interaction potential is convex, phase transitions of the system arise from nonconvexity features of the local vector field $F(x)$.

Model A was introduced in Ref. [29] to study the effect of multiplicative noise on spatially extended systems. We consider the Desai-Zwanzig model [6,35,52] settings where the local dynamics $F(x) = -V_\alpha'(x)$ is given by a double well potential $V_\alpha(x) = \frac{x^4}{4} - \alpha\frac{x^2}{2}$ and the noise is additive $\sigma(x) = \sigma$. The equations for motions are given by

$$dx_i = \left[\alpha x_i - x_i^3 - \theta(x_i - \bar{x})\right]dt + \sigma dW_i, \tag{A3}$$

where the Ito convention is now used. The above equations describe a system at equilibrium. In the $N \to \infty$ limit, it is useful to introduce the free energy functional $F[\rho]$ such that

$$\begin{aligned} F[\rho] &= \int dx V_\alpha(x)\rho(x) + \frac{\theta}{2}\iint dx dy \rho(x)\mathcal{U}(x-y)\rho(y) \\ &\quad + \frac{\sigma^2}{2}\int dx \rho(x)\ln\rho(x) \\ &:= \mathcal{V}[\rho] + \theta\mathcal{W}[\rho,\rho] - \frac{\sigma^2}{2}\mathcal{S}[\rho]. \end{aligned} \tag{A4}$$

The above equation describes energy balance in the system: $\mathcal{V}[\rho]$ represents the internal energy associated to the local potential $V_\alpha(x)$, $\mathcal{W}[\rho,\rho]$ is the energy given by the interaction among the agents and, lastly, $\mathcal{S}[\rho]$ is an entropic

contribution. As explained in the main text, the empirical measure $\rho_N = \frac{1}{N}\sum_i^N \delta_{x_i(t)}$ converges in the $N \to \infty$ limit to a one agent distribution $\rho(x,t)$ satisfying a nonlinear and nonlocal Fokker-Planck equation. The corresponding nonlinear Fokker-Planck equation of Eq. (A3) can be written in terms of the free energy as

$$\partial_t \rho = \frac{\partial}{\partial x}\left(\rho \frac{\partial}{\partial x}\frac{\delta F}{\delta \rho}\right). \tag{A5}$$

Remarkably, this equation belongs to a rich class of dissipative PDEs, including the heat equation, the porous medium equation and the diffusion-aggregation equation, that are gradient flows with respect to the Wasserstein metric on the space of probability measure with finite second moment, see Ref. [53] and references therein. The free energy $F[\rho]$ is a Lyapunov function for the dynamics and stationary solutions of the McKean-Vlasov equation are critical points of the free energy functional. In fact, the time derivative of $F[\rho]$ along solutions of Eq. (A5) is [53,54]

$$\frac{dF[\rho]}{dt} = -\int dy \rho(y)\left(\frac{\partial}{\partial y}\frac{\delta F}{\delta \rho}\right)^2 \leqslant 0 \tag{A6}$$

If an unique minimiser of the free energy exists, the dynamics converge exponentially fast, in relative entropy, to the unique stationary state and the rate of convergence to equilibrium can be established [55]. However, the minimiser is not necessarily unique and multiple stationary solutions can coexist. Furthermore, convexity properties of the free energy functional provide a one-to-one characterization of the stability properties of the stationary solutions. The model we have investigated in the main text arises from the assumption that the parameter α is not known exactly but rather erratically fluctuates in time, that is $\alpha \to \alpha + \sigma_m d\xi$ where $d\xi$ is another, uncorrelated, Brownian motion. This results in a set of equations for the N interacting agents that reads

$$dx_i = [-V'(x_i) - \theta(x_i - \bar{x})]\,dt + \sigma_m x_i \circ d\xi + \sigma dW_i \tag{A7}$$

where the symbol $\circ$ stands for a generic (not necessarily Ito) prescription for the equations. It is convenient to write the above set of equations in the equivalent, in law, form

$$dx_i = [-V'(x_i) - \theta(x_i - \bar{x})]\,dt + \sigma(x_i) \circ^{\nu} dW_i, \tag{A8}$$

where $\sigma(x) = \sqrt{\sigma^2 + \sigma_m^2 x^2}$ is a state-dependent stochastic term. It is well known that the presence of multiplicative noise introduce a modeling issue, since it is not clear, a priori, what prescription should be given to the stochastic integral defining the stochastic equation [56–58]; see also discussion in Ref. [48]. We interpret Eq. (A8) as a generic one parameter family of stochastic integrals parametrised by a parameter $\nu \in [0, 1]$. Different values of ν correspond to different prescription of the SDEs. In particular, $\alpha = 0, 1/2, 1$ correspond to the Ito, Stratonovich, and Klimontovich prescription, respectively. Different conventions of the stochastic integral lead to different stability properties of the SDE. Remarkably, the convention for a given system might also vary depending on the operational conditions [59]. In the main text of the paper, we always choose $\nu = \frac{1}{2}$. It is known that a generic SDE can be transformed into an Ito-SDE by suitably modifying the drift coefficient as $F_\alpha(x) \to F_{\alpha,\nu}(x) = F_\alpha(x) + \nu\sigma(x)\sigma'(x)$ [56]. Since it is more convenient to work with the Ito prescription, we apply this transformation to Eq. (A8) and obtain

$$dx_i = [-V_\nu(x_i) - \theta(x_i - \bar{x})]dt + \sigma(x_i)dW_i, \tag{A9}$$

where $V_{\alpha,\nu}(x) = V_\alpha(x) + \nu\sigma_m^2\frac{x^2}{2} = \frac{x^4}{4} - (\alpha + \nu\sigma_m^2)\frac{x^2}{2}$.

The introduction of a fluctuating parameter in the drift term corresponds to applying an external, state-dependent noise that breaks the detailed balance condition, thus driving the $N-$particle system to an out of equilibrium state. Equation (3) in the main text yields in this setting

$$f_{\langle x\rangle}(x) = -\frac{\alpha - \theta + (\nu - 1)\sigma_m^2 + \frac{\sigma^2}{\sigma_m^2}}{\sigma_m^2}\ln\left(1 + \left(\frac{\sigma_m}{\sigma}x\right)^2\right) + \frac{x^2}{\sigma_m^2} - 2\frac{\theta\langle x\rangle}{\sigma\sigma_m}\arctan\left(\frac{\sigma_m}{\sigma}x\right). \tag{A10}$$

The analysis of the self-consistency Eq. (5) (main text) provides insightful information on the stationary phase diagram of the model. In particular, symmetries of the problem force the system to always have the trivial solution $m^\star = 0$, corresponding to disordered state $\rho_0(x;0)$ of vanishing order parameter. This can be easily shown by observing that $R(-m) = -R(m)$ since stationary distributions satisfy $\rho_0(x;m) = \rho_0(-x;-m)$, see Eqs. (3) and (4) in the main text. Moreover, if $m^\star$ is a solution of the self-consistency equation, so is $-m^\star$. We thus expect that two symmetric branches of stable solutions will arise as soon as the disordered state loses stability. The disordered state becomes unstable as soon as $R'(0) = 1$, which reads

$$\frac{\theta}{\sigma\sigma_m}\left\langle x \arctan\left(\frac{\sigma_m}{\sigma}x\right)\right\rangle_0 = \frac{1}{2}, \tag{A11}$$

where the expectation value $\langle\cdot\rangle_0$ is taken with respect to the stationary distribution $\rho_0(x;0)$. Since the order parameter vanishes at the transition point, the above equation yields, fixed all the other parameters, the critical value $\sigma_c = \sigma_c(\alpha, \theta, \sigma_m)$ of the strength of the additive noise. Figure 3 shows the multiplicative noise induced stabilization phenomenon we mentioned in the main text. Indeed, the multiplicative noise has a rectifying effect, pushing, for strong enough coupling θ, the transition point to higher and higher values of σ. Moreover, the amplitude of the order parameter gets magnified, since it exceeds the maximum value $\sqrt{\alpha}$, the minimum point of the potential $V_\alpha(x)$, that is attained in the low noise regime ($\sigma \to 0$) when $\sigma_m = 0$.

Model B features a discontinuous phase transition and is obtained by breaking the symmetry $x \to -x$ through a *tilted* potential as $V_{\alpha,k} = V_\alpha + \mu x$, with $\mu > 0$. Moreover, the system is subject to thermal noise $\sigma(x) = \sigma$. The pitchfork bifurcation of invariant solutions one obtains for $\mu = 0$ disappears. In particular, there exists a smooth, stable branch of negative order parameter $\langle x\rangle$ for all values of the strength of the noise σ. However, decreasing σ, a pair of solutions appear through a saddle node bifurcation, yielding another branch of stable $\langle x\rangle > 0$, with the other one being unstable, see Fig. 1 in the main text. The saddle node bifurcation is characterised by

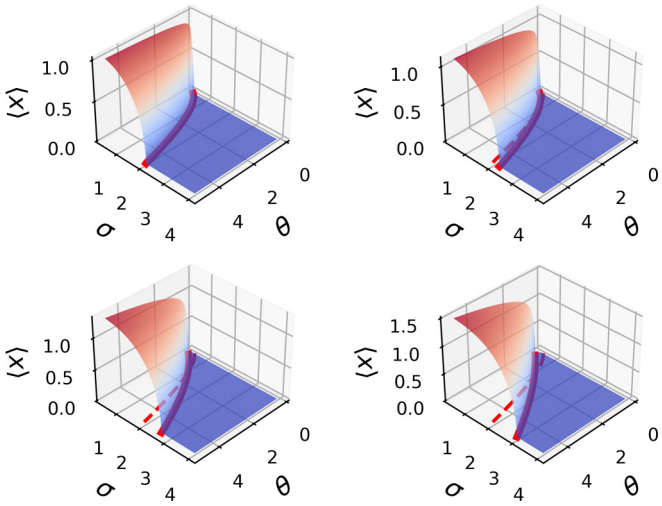


FIG. 3. Order parameter $\langle x\rangle$ as a function of (σ,θ) obtained via the self-consistency equation analysis. The red dashed line and the continuous red line represent the exact transition curve for $\sigma_m = 0$ and $\sigma_m \neq 0$, see Eq. (A11). The other parameters of the model are fixed and equal to $\alpha = 1$, $\sigma_m = 1.5$, $\nu = 1/2$.

the condition $R'(m_c) = 1$ that reads

$$\frac{\theta}{\sigma^2}\langle (x-m_c)^2\rangle_0 = \frac{1}{2}, \tag{A12}$$

where m_c is the value of the positive order parameter at the transition point and the expectation value is taken with respect to the stationary distribution $\rho_0(x;m_c)$. Since m_c is not known a priori and has to be evaluated numerically by solving the self-consistency equation, the above equation does not directly provide the value of the critical noise σ_c at which the saddle node bifurcation takes place. Nevertheless, it provides a criterion to assess how close the critical point evaluated numerically is to the exact one by evaluating the slope $R'(m_c)$ and comparing it to the exact value 1.

Model B' interesting feature is represented by the discontinuous phase transition and the jump from the top branch to the bottom one as the parameter σ is changed. Such analysis has been performed in the main text. Nevertheless one could study the dynamical response of the system as the transition point is approached from below on the top branch. Since it is associated with the loss of stability of the invariant measure, we expect similar results to hold for this model as well. We refer to the main text and to Appendix D for the explanation (and for the notations) of the linear response investigation we have performed. Figure 4 shows that the Green function associated to the order parameter $\langle x\rangle$ and a time delta $\delta(t)$ homogeneous perturbation develops a timescale, for settings near the phase transition, that is orders of magnitude bigger than the timescale associated to noncritical settings. We remark that such behavior does not depend on the specific form of the forcing [13]. The figure refers to a level of truncation of $\bar{n} = 22$. One could also perform an analysis by looking at different values of $\bar{n}$. We expect to obtain similar results to what is reported in the main text. However, such analysis is more complicated here by the discontinuous feature of the transition. Firstly, the reduced dynamics transition point depends on $\bar{n}$ and the analysis becomes increasingly hard very close to the transition point, see shaded area in panel (b) of Fig. 1 in the main text. Secondly, Fig. 4, clearly shows that the timescale associated to the Green function is highly sensitive to small deviations, such as $\delta = 0.1\%$, from the transition point.

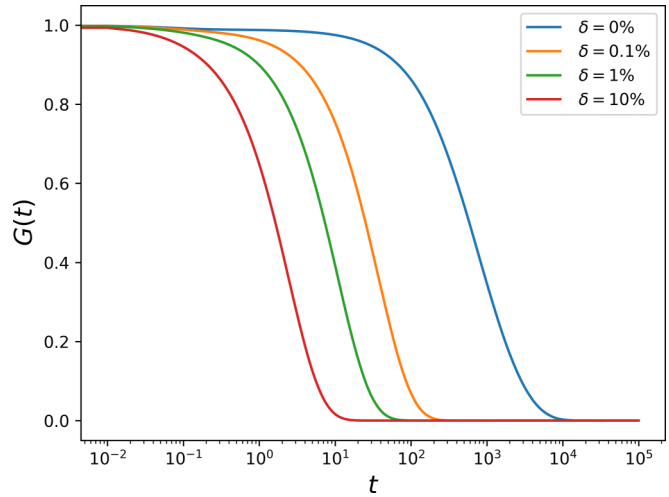


FIG. 4. Green function $G(t)$ as a function of time for model B. δ represents the relative distance from the phase transition point.

APPENDIX B: HIERARCHY OF EQUATIONS FOR THE MOMENTS AND CUMULANTS

In this section we provide a few more details on how to obtain the dynamical evolution of the moments and cumulants of the distribution of the infinite system $\rho(x,t)$. As explained in the main text, $\rho(x,t)$ satisfies a nonlinear and nonlocal Fokker-Planck equation that we write here in an alternative way as

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial x}((\hat{F}_\alpha(x) + \theta(x-\langle x\rangle))\rho) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(x)\rho), \tag{B1}$$

where $\hat{F}_\alpha = F_\alpha + \frac{1}{2}\sigma(x)\sigma'(x)$. If we multiply (B1) by x^n and integrate on the phase space $\mathbb{R}$, we obtain after performing some integration by parts

$$\begin{aligned}\frac{dM_n}{dt} &= n(\langle \hat{F}_\alpha x^{n-1}\rangle - \theta\langle (x-\langle x\rangle)x^{n-1}\rangle) \\ &\quad + \frac{n(n-1)}{2}\langle x^{n-2}\sigma^2(x)\rangle \\ &= n(\langle \hat{F}_\alpha x^{n-1}\rangle - \theta(M_n - M_1 M_{n-1})) \\ &\quad + \frac{n(n-1)}{2}\langle x^{n-2}\sigma^2(x)\rangle,\end{aligned} \tag{B2}$$

where $\langle\cdot\rangle$ represents the expectation value with respect to the probability distribution ρ and we have introduced the moments $M_n = \langle x^n\rangle$. We observe that the main assumption in this paper, namely the fact that we assume that the local drift F_α and the diffusion coefficient $\sigma^2(x)$ have a polynomial functional form, implies that both $\langle \hat{F}_\alpha x^{n-1}\rangle$ and $\langle x^{n-2}\sigma^2(x)\rangle$ can be written in a closed form in terms of the moments M_n. Indeed, let us explicitly carry out these calculations for model

A. Similar results hold for model B. We recall that model A is defined by a diffusion coefficient is $\sigma^2(x) = \sigma^2 + \sigma_m^2 x^2$ and a local drift $F_\alpha(x) = \alpha x - x^3$, hence

$$\begin{aligned}\langle x^{n-2}\sigma^2\rangle &= \langle x^{n-2}(\sigma^2 + \sigma_m^2 x^2)\rangle \\ &= \sigma^2 M_{n-2} + \sigma_m^2 M_n, \\ \langle \hat{F}_\alpha x^{n-1}\rangle &= \langle (F_\alpha + \tfrac{1}{2}\sigma^2 x)x^{n-1}\rangle \\ &= \alpha M_n - M_{n+2} + \tfrac{1}{2}\sigma_m^2 M_n.\end{aligned} \tag{B3}$$

From (B2) one then obtains an infinite hierarchy of equations for the moments as

$$\begin{aligned}\frac{dM_n}{dt} =& n\left(\alpha - \theta + n\frac{\sigma_m^2}{2}\right)M_n - M_{n+2} \\ &+ \frac{n(n-1)}{2}\sigma^2 M_{n-2} + \theta M_1 M_{n-1}.\end{aligned} \tag{B4}$$

The above calculations have been obtained for a quadratic interaction potential $\mathcal{U}(x) = \frac{x^2}{2}$, but we remark that infinite hierarchies of equations for the moments such as (B4) can be obtained for any polynomial interaction potential $\mathcal{U}(x)$. If the functions describing the dynamics are generic, as opposed to polynomials, it is not possible to find close equations for the moments. However, one could potentially recur to a Taylor expansion to approximate, in a controlled way, these functions as polynomials and then construct the corresponding approximate hierarchy of equations for the moments. Of course, this would introduce another source of approximation on top of the one deriving from the truncation scheme of the hierarchy.

Following Ref. [35], one can alternatively obtain an infinite hierarchy of equations for the cumulants of the probability distribution ρ. We remark that the cumulants k_n are defined through the cumulant generating function $G(\lambda, t) = \ln g(\lambda, t)$ as

$$\sum_{n=1}^{\infty} k_n(t)\frac{\lambda^n}{n!} = \ln \int \rho(x,t)e^{\lambda x}dx \equiv \ln g(\lambda, t). \tag{B5}$$

Equation (B1) yields an evolution equation for the cumulant generating function

$$\begin{aligned}\frac{dG}{dt} =& \frac{1}{g}\frac{dg}{dt} = \frac{1}{g}\int \frac{\partial \rho}{\partial t}e^{\lambda x}dx \\ =& -\frac{\lambda}{g}\int dx(x^3 - (\alpha - \theta + \nu\sigma^2 x^2) - \theta\langle x\rangle)\rho e^{\lambda x} \\ &+ \frac{\lambda^2}{2g}\int dx(\sigma^2 + \sigma_m^2 x^2)\rho e^{\lambda x}.\end{aligned} \tag{B6}$$

By separating the different powers of the variable x we can write the above equation in terms of G, its derivative $G'(\lambda, t) = \frac{\partial G}{\partial \lambda}$ and higher order derivatives as

$$\begin{aligned}\frac{dG}{dt} =& \lambda\theta\langle x\rangle + \frac{\lambda^2\sigma^2}{2} + \lambda(\alpha - \theta + \nu\sigma^2)G' \\ &+ \frac{\lambda^2\sigma_m^2}{2}(G'^2 + G'') \\ &- \lambda(G'G'^2 + 3G'G'' + G''').\end{aligned} \tag{B7}$$

Using the definition of the cumulants given in Eq. (B5) and comparing same powers of λ one finally obtains the equations for the cumulants

$$\begin{aligned}\frac{1}{n}\frac{dk_n}{dt} =& \theta k_1\delta_{1n} + \frac{\sigma^2}{2}\delta_{n2} \\ &+ \left(\alpha - \theta + \sigma_m^2\left(\nu + \frac{n-1}{2}\right)\right)k_n - k_{n+2} \\ &+ \sigma_m^2(1-\delta_{n1})\frac{(n-1)!}{2}\sum_{i=1}^{n-1}\frac{k_i k_{n-i}}{(i-1)!(n-i-1)!} \\ &- 3(n-1)!\sum_{i=1}^{n}\frac{k_i k_{n-i+2}}{(i-1)!(n-i)!} - (n-1)!\sum_{i=1}^{n} \\ &\times \sum_{j=1}^{n-i+1}\frac{k_i k_j k_{n+2-i-j}}{(i-1)!(j-1)!(n-i-j+1)!}.\end{aligned} \tag{B8}$$

APPENDIX C: TRUNCATION SCHEMES

This section is divided in two parts. In the first, we provide the algebra to perform a cumulant truncation scheme at any generic order n for the hierarchy of equations for the moments (B4). Secondly, we compare the performances of multiple truncation schemes and assess that the cumulant truncation scheme correspond to the best parametrization choice for the thermodynamic limit of the interacting agents system.

1. Cumulant truncation scheme

The relationship between cumulants and moments of a probability distribution is

$$k_n = \sum_{l=1}^{n}(-1)^{l-1}(l-1)!B_{nl}(M_1, \ldots, M_{n-l+1}), \tag{C1}$$

where $B_{nl}(M_1, \ldots, M_{n-l+1})$ are partial (incomplete) Bell polynomials. In particular, these polynomials are given by

$$\begin{aligned}B_{nl}(M_1, \ldots, M_{n-l+1}) =& \sum \frac{n!}{j_1!j_2!\ldots j_{n-l+1}!}\left(\frac{M_1}{1!}\right)^{j_1} \\ &\times \left(\frac{M_2}{2!}\right)^{j_2}\cdots\left(\frac{M_{n-l+1}}{(n-l+1)!}\right)^{j_1},\end{aligned} \tag{C2}$$

where the sum is taken over all the sequences $j_1 j_2 \ldots j_{n-l+1}$ of non-negative integers such that the following two conditions hold:

$$\begin{aligned}&j_1 + j_2 + \ldots j_{n-l+1} = l, \\ &j_1 + 2j_2 + \cdots + (n-l+1)j_{n-l+1} = n.\end{aligned}$$

Moreover, we will make extensive use of the following two properties of the Bell polynomials:

$$B_{n1}(M_1, \ldots, M_n) = M_n, \tag{C3}$$

$$B_{n2}(M_1, \ldots, M_{n-1}) = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k}M_k M_{n-k}. \tag{C4}$$

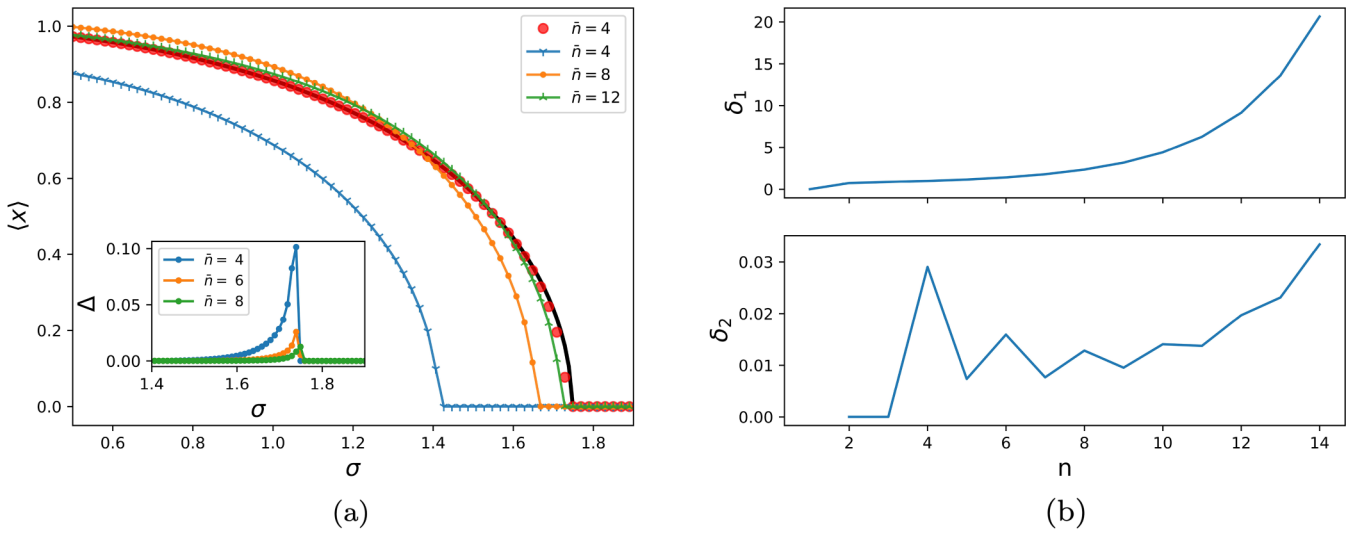


FIG. 5. (a) Phase diagram for model A. The continuous black line corresponds to the phase diagram as obtained from the self-consistency equation, see main text. The red dots correspond to the CT scheme of order $\bar{n} = 4$. The continuous lines with markers correspond instead to a MT schemes of increasing order. The bottom left inset shows the absolute error Δ between the self-consistency equation and the CT scheme. (b) The top (bottom) panel shows the difference in magnitude between moments (central moments) and cumulants for increasing order of truncation. Moments, central moments, and cumulants have been obtained from the known expression of the invariant distribution $\rho_0(x; m)$, where m has been evaluated through the self-consistency equation, see main text. Here the parameters are $(\alpha, \theta, \sigma_m, \nu) = (1, 4, 0.2, 0.5)$. Moreover, in (b), $\sigma \approx 1$.

The closure approximation $\bar{M}_{\bar{n}+1}$ can be easily found by separating the term $l = 1$ from Eqs. (C1) and using (C3),

$$k_n = M_n + \sum_{l=2}^{n} (-1)^{l-1}(l-1)!B_{nl}(M_1, \ldots, M_{n-l+1}). \quad \text{(C5)}$$

In fact, evaluating the above equation for $n = \bar{n} + 1$ and imposing the condition $k_{\bar{n}+1} = 0$ results in

$$\bar{M}_{\bar{n}+1} = -\sum_{l=2}^{\bar{n}+1} (-1)^{l-1}(l-1)!B_{\bar{n}+1,l}(M_1, \ldots, M_{\bar{n}+2-l}). \quad \text{(C6)}$$

The evaluation of $\bar{M}_{\bar{n}+2}$ requires more care since it involves $\bar{M}_{\bar{n}+1}$ as well. Let us first observe that the cumulant $k_{\bar{n}+2}$ can be written as, see Eq. (C1),

$$\begin{aligned} k_{\bar{n}+2} =& M_{\bar{n}+2} - B_{\bar{n}+2,2}(M_1, \ldots, M_{\bar{n}+1}) \\ &+ \sum_{l=1}^{n} (-1)^{l-1}(l-1)!B_{\bar{n}+2,l}(M_1, \ldots, M_{\bar{n}+3-l}). \end{aligned} \quad \text{(C7)}$$

Using Eq. (C4), we can write

$$\begin{aligned} B_{\bar{n}+2,2}(M_1, \ldots, M_{\bar{n}+1}) =& (\bar{n}+2)M_{\bar{n}+1}M_1 \\ &+ \sum_{k=2}^{\bar{n}} \binom{\bar{n}+2}{k} M_k M_{\bar{n}+2-k}, \end{aligned} \quad \text{(C8)}$$

where we have separated the term $k = 1$ and $k = \bar{n} + 1$ from the total sum.

Finally, by imposing the condition $k_{\bar{n}+2} = 0$ and consistently estimating $M_{\bar{n}+1}$ as $\bar{M}_{\bar{n}+1}$ we obtain the approximated value for $M_{\bar{n}+2}$ as

$$\begin{aligned} \bar{M}_{\bar{n}+2} =& (\bar{n}+2)\bar{M}_{\bar{n}+1}M_1 + \frac{1}{2}\sum_{k=2}^{\bar{n}} \binom{\bar{n}+2}{k} M_k M_{\bar{n}+2-k} \\ &- \sum_{l=3}^{\bar{n}+2} (-1)^{(l-1)}(l-1)!B_{\bar{n}+2,l}(M_1, \ldots, M_{\bar{n}+3-l}). \end{aligned} \quad \text{(C9)}$$

In conclusion, the cumulant truncation scheme consists in the finite set of Eq. (6) with $n = 1, \ldots, \bar{n}$ along with the boundary conditions $M_0 = 1$ and $M_{\bar{n}+1} = \bar{M}_{\bar{n}+1}$, $M_{\bar{n}+2} = \bar{M}_{\bar{n}+2}$ as given by Eqs. (C6) and (C9), respectively.

2. Comparison between different truncation schemes

The infinite hierarchy of equation for the moments (B4) or cumulants (B8) are equivalent to the McKean Vlasov equation (B1) describing the thermodynamic limit of the interacting agents system. For obvious practical reasons, it is necessary to find appropriate truncation schemes to the hierarchy resulting in a finite, preferably small, number of ordinary differential equations for the moments or cumulants. In particular, common truncation schemes include a moment truncation scheme (MT), a central moment truncation scheme (cMT) and a cumulant truncation scheme (CT). These schemes correspond to imposing ad hoc boundary conditions to the hierarchy of moments or cumulants. Following Refs. [35,40], we have implemented in the main text the CT scheme and proved that the cumulants act as effective reaction coordinates for the system. The low-dimensional reduced order dynamics for a small number of cumulants, resulting from the CT scheme, is able to capture both stationary and time-dependent properties of the thermodynamic limit of the

interacting agents system. We recall that the CT scheme of order $\bar{n}$ is equivalent to imposing the condition $k_{\bar{n}+1} = k_{\bar{n}+2} = 0$ in Eq. (B8). This is equivalent, as explained in the previous section, to imposing the boundary conditions (C6) and (C9) to the hierarchy of equations for the moments (B4). Instead, the MT scheme at level $\bar{n}$ is obtained by imposing the condition $M_{\bar{n}+1} = M_{\bar{n}+2} = 0$ for Eq. (B4). Similarly, when the above vanishing condition is applied to the central moments one obtains the cMT scheme. Figure 5 provides a quantitative comparison between the three approaches and clarifies why the CT is preferable in our settings. Panel (a) shows the phase diagram of the system. The black solid line derives from solving numerically the self-consistency equation and provides a reference point for the approximate results stemming from the reduced dynamics obtained from the CT (red dots) and the MT (lines with markers) schemes. It is clear that a parametrization in terms of cumulants provides a better approximation, fixed the order $\bar{n}$, of the dynamics of the system than a parametrization in terms of moments. As shown in the main text too, a parametrization in terms of as low as $\bar{n} = 4$ cumulants yields a good approximation of the stationary dynamics, see also the bottom left inset showing the absolute error Δ between the CT and the self-consistency equation. In particular, as explained in the main text, near the phase transition point one needs to include a higher number of reaction coordinates to achieve a better performance. On the contrary, the MT scheme yields a reduced order dynamics that does not capture the stationary properties of the system in most of the range of values spanned by the strength of the noise σ.

In order to investigate in a quantitative way the difference between the three truncation schemes we introduce the metrics $\delta_1 = |M_n| - |k_n|$ and $\delta_2 = |M'_n| - |k_n|$, where we have denoted with M'_n the central moment of order n. These metrics provide a measure, at each order of truncation n, of the difference of the magnitudes of the moments and central moments with respect to the corresponding cumulant. Panel (b) shows that δ_1 and δ_2 are positive meaning that the cumulants k_n are, in magnitude, always smaller than the corresponding (central) moments, validating *a posteriori* our choice of using a CT scheme.

APPENDIX D: LINEAR RESPONSE THEORY FOR MCKEAN-VLASOV EQUATION: SINGULARITIES OF THE SUSCEPTIBILITY

In this section we provide more details about the linear response properties of model A. The ultimate goal of this section is to prove the formula for the residue of the singular part of the susceptibility $\chi(\omega)$ at the phase transition.

The invariant measures $\rho_0(x)$ of the McKean-Vlasov equation, see Eq. (2) in the main text, satisfy the eigenvalue problem $\mathcal{L}_{\langle x\rangle_0}\rho_0(x) = 0$, where the linear differential operator $\mathcal{L}_{\langle x\rangle_0}$ is defined by

$$\mathcal{L}_{\langle x\rangle_0}\psi(x) = \frac{\partial}{\partial x}\left(\frac{\sigma^2(x)}{2}\psi\frac{\partial}{\partial x}(f_{\langle x\rangle_0}(x) + \ln\psi)\right), \tag{D1}$$

where $\psi(x)$ is a smooth function and $f_{\langle x\rangle_0}(x)$ is defined in Eq. (3) in the main text. We now perturb the stationary state by applying a perturbation to the drift $F_\alpha(x) \to F_\alpha(x) + \varepsilon X(x)T(t)$, where $\varepsilon \ll 1$. We can observe the effect of the perturbation in terms of the measure of the system as $\rho(x,t) = \rho_0(x) + \varepsilon\rho_1(x,t) + \dots$ Alternatively, we can investigate the time-dependent properties of any observable of the system after the perturbation. In the following, we will observe the response of the order parameter $\langle x\rangle$ and write $\langle x\rangle = \langle x\rangle_0 + \varepsilon\langle x\rangle_1(t)$ where $\langle\cdot\rangle_1$ represents the expectation value with respect to the measure $\rho_1(x,t)$. We define the Fourier transform of any function $f(t)$ as $f(\omega) = \int f(t)e^{i\omega t}dt$. The response of the order parameter in frequency space is given by [13,14]

$$\langle x\rangle_1(\omega) = \chi(\omega)T(\omega), \tag{D2}$$

where the susceptibility $\chi(\omega)$ is written as

$$\chi(\omega) = \frac{\Gamma(\omega)}{1 - \theta\Gamma(\omega)}. \tag{D3}$$

The microscopic susceptibility $\Gamma(\omega)$ is related to microscopic correlation properties of the system in the unperturbed state described by ρ_0. In particular, $\Gamma(\omega)$ is the Fourier transform of the microscopic response function $\Gamma(t)$ that can be written as a suitable correlation function as [13]

$$\Gamma(t) = -\Theta(t)\left\langle\frac{1}{\rho_0(x)}\frac{\partial}{\partial x}(\rho_0 X(x))\exp(\mathcal{L}^\dagger_{\langle x\rangle_0}t)x\right\rangle_0, \tag{D4}$$

where the operator $\mathcal{L}^\dagger_{\langle x\rangle_0}$ is the adjoint of $\mathcal{L}_{\langle x\rangle_0}$ and can be interpreted as the generator of the Koopman operator of the stationary dynamics described by $\rho_0(x)$. For gradient systems with thermal noise, it is possible to write $\Gamma(t)$ as a time derivative of suitable correlation properties. We remark that for general nonequilibrium systems this is not always possible. However, given the structure of the problem, we are able find an analogous formula for $\Gamma(t)$. As described in the main text, we evaluate the response of the system to a homogeneous perturbation $X(x) = 1$. The microscopic response function, see Eq. (D4), is

$$\begin{aligned}\Gamma(t) &= -\Theta(t)\int dx\frac{\partial\rho_0}{\partial x}\exp(\mathcal{L}^\dagger_{\langle x\rangle_0}t)x\\ &= -\Theta(t)\int dx x\exp(\mathcal{L}_{\langle x\rangle_0}t)\frac{\partial\rho_0}{\partial x}\\ &= +\Theta(t)\int dx x\exp(\mathcal{L}_{\langle x\rangle_0}t)\rho_0(x)\frac{\partial}{\partial x}f_{\langle x\rangle_0}(x),\end{aligned} \tag{D5}$$

where we have used the definition of the adjoint of an operator and Eq. (4) in the main text to evaluate the derivative of the stationary distribution. We now define the function $g(x) = -\frac{1}{\sigma\sigma_m}\arctan(\frac{\sigma_m}{\sigma}x)$ such that its derivative is $\frac{\partial g(x)}{\partial x} = -\frac{1}{\sigma^2(x)}$. We then evaluate the following expression:

$$\begin{aligned}\mathcal{L}_{\langle x\rangle_0}(g\rho_0) &= \frac{\partial}{\partial x}\left(\frac{\sigma(x)^2}{2}g\rho_0\frac{\partial}{\partial x}\left(f_{\langle x\rangle_0}(x) + \ln\rho_0 + \ln g\right)\right)\\ &= \frac{\partial}{\partial x}\left(\frac{\sigma(x)^2}{2}g\rho_0\frac{\partial}{\partial x}\ln g\right) = \frac{\partial}{\partial x}\left(\frac{\sigma(x)^2}{2}\rho_0\frac{\partial}{\partial x}g\right)\\ &= -\frac{1}{2}\frac{\partial}{\partial x}\rho_0 = +\frac{1}{2}\rho\frac{\partial}{\partial x}f_{\langle x\rangle_0}(x),\end{aligned} \tag{D6}$$

where we have used the fact that $f_{\langle x\rangle_0}(x) + \ln\rho_0 = Z =$ constant.

The microscopic response function can thus be written as

$$\begin{aligned}\Gamma(t) &= -\frac{2}{\sigma\sigma_m}\Theta(t)\int dx x \exp(\mathcal{L}_{\langle x\rangle_0}t)\mathcal{L}_{\langle x\rangle_0}\arctan\left(\frac{\sigma_m}{\sigma}x\right)\rho_0(x)\\ &= -\frac{2}{\sigma\sigma_m}\Theta(t)\frac{d}{dt}\int dx x \exp(\mathcal{L}_{\langle x\rangle_0}t)\arctan\left(\frac{\sigma_m}{\sigma}x\right)\rho_0(x)\\ &= -\frac{2}{\sigma\sigma_m}\Theta(t)\frac{d}{dt}C_{x,A}(t), \end{aligned} \tag{D7}$$

where in the last line we have introduced the correlation function between observable x and observable $A = \arctan(\frac{\sigma_m}{\sigma}x)$ defined as

$$\begin{aligned}C_{x,A}(t) &= \langle x(t)A(x(0))\rangle_0 - \langle x\rangle_0\langle A\rangle_0\\ &= \int dx x \exp(\mathcal{L}_{\langle x\rangle_0}t)A(x)\rho_0(x) - \langle x\rangle_0\langle A\rangle_0. \end{aligned} \tag{D8}$$

The microscopic susceptibility can thus be written as

$$\Gamma(\omega) = \int_{-\infty}^{+\infty} dt e^{i\omega t}\Gamma(t) = \frac{2}{\sigma\sigma_m}(C_{x,A}(0) + i\omega\hat{C}_{x,A}(\omega)), \tag{D9}$$

where $\hat{C}_{x,A}(\omega) = \int_0^{+\infty} e^{i\omega t}C_{x,A}(t)$ is the (one-sided) Fourier transform of the correlation function $C_{x,A}(t)$.

We can then show that the macroscopic susceptibility $\chi(\omega)$ develops a singular behavior for a real frequency $\omega_0 = 0$ at the phase transition. Let us observe that Eq. (A11), that characterises the phase transition line, can be written as

$$\frac{\theta}{\sigma\sigma_m}C_{x,A}(0) = \frac{1}{2} \tag{D10}$$

since $\langle x\rangle_0 = 0$ at the transition point. In conclusion, using all the above results, the susceptibility $\chi(\omega)$ of the system is

$$\chi(\omega) = -\frac{1}{\theta} + i\frac{1}{\omega}\frac{\sigma\sigma_m}{\theta^2\hat{C}_{x,A}(\omega)} = -\frac{1}{\theta} + i\frac{1}{\omega}\frac{C_{x,A}(0)}{\theta\hat{C}_{x,A}(\omega)}. \tag{D11}$$

Being related to the spectral properties of the operator $\mathcal{L}_{\langle x\rangle_0}$, the quantity $\hat{C}_{x,A}(\omega)$ is an analytical function at the phase transition [13,14,52]. Consequently, the above equation shows that linear response theory breaks down at the phase transition, with the susceptibility $\chi(\omega)$ developing a simple pole in $\omega = \omega_0 = 0$ with residue

$$\kappa = \operatorname*{Res}_{\omega=\omega_0}\chi(\omega) = \frac{i}{\theta}\frac{C_{x,A}(0)}{\hat{C}_{x,A}(0)} = \frac{i}{\theta\tau_{x,A}}, \tag{D12}$$

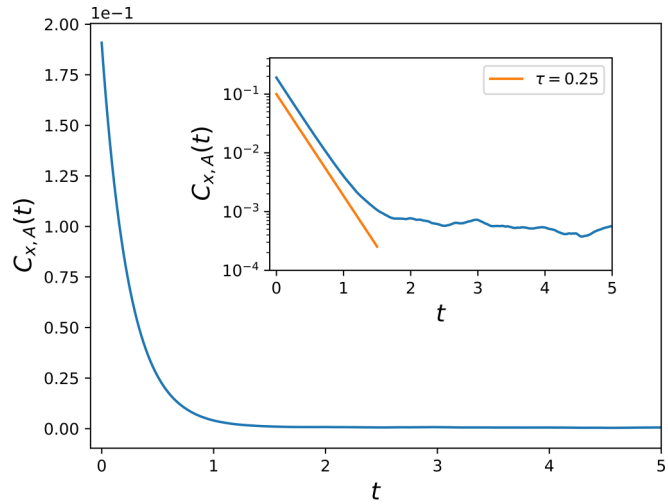


FIG. 6. Correlation function $C_{x,A}(t)$ as a function of time. The orange line in the inset corresponds to an exponentially decaying function $y = 0.1e^{-t/\tau}$ where $\tau = 0.25$. The parameters of the model are the same as in Fig. 2 of the main text.

where $\tau_{x,A}$ is the integrated auto-correlation time defined by

$$\tau_{x,A} = \frac{\hat{C}_{x,A}(0)}{C_{x,A}(0)} = \frac{\int_0^{+\infty} C_{x,A}(t)dt}{C_{x,A}(0)}. \tag{D13}$$

As $\sigma_m \to 0$, the above equations are compatible with the results of Ref. [52]. We have numerically estimated the correlation function $C_{x,A}(t)$ by evaluating the one-agent correlation function $c_i(t)$ between x_i and $A(x_i)$ and then averaging over the whole ensemble of agents ($N = 16000$), thus yielding $C_{x,A}(t) = \frac{1}{N}\sum_{i=1}^{N} c_i(t)$. The integrated correlation time $\tau_{x,A}$ has been estimated by imposing a cut off $T = 1.5$ on the time integral corresponding to the moment after which the noisy signal takes over the exponential decay of the correlation function (see inset of Fig. 6). The resulting value is $\tau = 0.25091$ with corresponding amplitude of the residue $k = 0.99636$, which agrees with what has been obtained through the reduced order dynamics, see Fig. 2 in the main text.

[1] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials (Springer Cham, 2016).

[2] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Multilayer networks, Journal of Complex Networks **2**, 203 (2014).

[3] S. Yanchuk, A. C. Roque, E. E. N. Macau, and J. Kurths, Dynamical phenomena in complex networks: Fundamentals and applications, Eur. Phys. J.: Spec. Top. **230**, 2711 (2021).

[4] G. Naldi, L. Pareschi, and G. Toscani, *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences* (Birkhäuser Basel, 2010).

[5] L. Pareschi and G. Toscani, *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods* (Oxford University Press, Oxford, 2014).

[6] D. A. Dawson, Critical dynamics and fluctuations for a mean-field model of cooperative behavior, J. Stat. Phys. **31**, 29 (1983).

[7] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, The kuramoto model: A simple paradigm for synchronization phenomena, Rev. Mod. Phys. **77**, 137 (2005).

[8] J. Garnier, G. Papanicolaou, and T. Yang, Large deviations for a mean-field model of systemic risk, SIAM Journal on Financial Mathematics **4**, 151 (2013).

[9] J. Garnier, G. Papanicolaou, and T. Yang, Consensus convergence with stochastic effects, Vietnam Journal of Mathematics **45**, 51 (2017).

[10] G. M. Rotskoff and E. Vanden-Eijnden, Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error, 2018.

[11] A. Garbuno-Inigo, N. Nüsken, and S. Reich, Affine invariant interacting Langevin dynamics for Bayesian inference, SIAM J. Appl. Dyn. Syst. **19**, 1633 (2020).

[12] A. Borovykh, N. Kantas, P. Parpas, and G. A. Pavliotis, On stochastic mirror descent with interacting particles: Convergence properties and variance reduction, Physica D **418**, 132844 (2021).

[13] V. Lucarini, G. A. Pavliotis, and N. Zagli, Response theory and phase transitions for the thermodynamic limit of interacting identical systems, Proc. R. Soc. A. **476**, 20200688 (2020).

[14] N. Zagli, V. Lucarini, and G. A. Pavliotis, Spectroscopy of phase transitions for multiagent systems, Chaos: An Interdisciplinary Journal of Nonlinear Science **31**, 061103 (2021).

[15] A. Ma and A. R. Dinner, Automatic method for identifying reaction coordinates in complex systems, J. Phys. Chem. B **109**, 6769 (2005).

[16] G. Bussi, A. Laio, and M. Parrinello, Equilibrium Free Energies from Nonequilibrium Metadynamics, Phys. Rev. Lett. **96**, 090601 (2006).

[17] J. Rogal, Reaction coordinates in complex systems-a perspective, Eur. Phys. J. B **94**, 223 (2021).

[18] L. L. Bonilla, J. Casado, and M. Morillo, Self-synchronization of populations of nonlinear oscillators in the thermodynamic limit, J. Stat. Phys. **48**, 571 (1987).

[19] A. Pikovsky, J. Kurths, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge Nonlinear Science Series, Cambridge University Press, 2003).

[20] F. Collet, P. Dai Pra, and M. Formentin, Collective periodicity in mean-field models of cooperative behavior, Nonlinear Differential Equations and Applications NoDEA **22**, 1461 (2015).

[21] P. Dai Pra, Stochastic mean-field dynamics and applications to life sciences, *Stochastic Dynamics Out of Equilibrium*, edited by G. Giacomin, S. Olla, E. Saada, H. Spohn, and G. Stoltz (Springer International Publishing, 2019), pp. 3–27.

[22] H. Mori, Transport, collective motion, and Brownian motion, Prog. Theor. Phys. **33**, 423 (1965).

[23] R. Zwanzig, Memory effects in irreversible thermodynamics, Phys. Rev. **124**, 983 (1961).

[24] J. Wouters and V. Lucarini, Disentangling multi-level systems: averaging, correlations and memory, J. Stat. Mech.: Theory Exp. (2012) P03003.

[25] J. Wouters and V. Lucarini, Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach, J. Stat. Phys. **151**, 850 (2013).

[26] M. D. Chekroun, H. Liu, and S. Wang, *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations*, Springer Briefs in Mathematics (Springer Cham, 2015).

[27] D. Kondrashov, M. D. Chekroun, and M. Ghil, Data-driven non-Markovian closure models, Physica D **297**, 33 (2015).

[28] M. Santos Gutiérrez, V. Lucarini, M. D. Chekroun, and M. Ghil, Reduced-order models for coupled dynamical systems: Data-driven methods and the koopman operator, Chaos: An Interdisciplinary Journal of Nonlinear Science **31**, 053116 (2021).

[29] C. Van den Broeck, J. M. R. Parrondo, J. Armero, and A. Hernández-Machado, Mean field model for spatially extended systems in the presence of multiplicative noise, Phys. Rev. E **49**, 2639 (1994).

[30] S. N. Gomes, S. Kalliadasis, G. A. Pavliotis, and P. Yatsyshin, Dynamics of the desai-zwanzig model in multiwell and random energy landscapes, Phys. Rev. E **99**, 032109 (2019).

[31] D. A. Dawsont and J. Gärtner, Large deviations from the mckean-vlasov limit for weakly interacting diffusions, Stochastics **20**, 247 (1987).

[32] A. Sznitman, *Topics in Propagation of Chaos.*, edited by P.L. Hennequin, Ecole d'Eté de Probabilités de Saint-Flour XIX—1989. Lecture Notes in Mathematics (Springer, Berlin, Heidelberg, 1989), Vol. 1464.

[33] K. Oelschlager, A martingale approach to the law of large numbers for weakly interacting stochastic processes, Ann. Probab. **12**, 458 (1984).

[34] S. Gomes and G. Pavliotis, Mean field limits for interacting diffusions in a two-scale potential, J. Nonlin. Sci. **28**, 905 (2018).

[35] R. C. Desai and R. Zwanzig, Statistical mechanics of a nonlinear stochastic model, J. Stat. Phys. **19**, 1 (1978).

[36] L. Fialkow, The truncated k-moment problem: A survey, *Operator Theory: The State of the Art*, Theta Ser. Adv. Math., Vol. 18 (Theta, Bucharest, Romania, 2016), pp. 25–51.

[37] M. Infusino, T. Kuna, J. Lebowitz, and E. Speer, The truncated moment problem on n0, J. Math. Anal. Appl. **452**, 443 (2017).

[38] F. J. Alexander, G. Johnson, G. L. Eyink, and I. G. Kevrekidis, Equation-free implementation of statistical moment closures, Phys. Rev. E **77**, 026701 (2008).

[39] L. H. Chan, K. Chen, C. Li, C. W. Wong, and C. Y. Yau, On higher-order moment and cumulant estimation, J. Stat. Comput. Simul. **90**, 747 (2020).

[40] R. M. Wilcox and R. Bellman, Truncation and preservation of moment properties for fokker-planck moment equations, J. Math. Anal. Appl. **32**, 532 (1970).

[41] R. Schack and A. Schenzle, Moment hierarchies and cumulants in quantum optics, Phys. Rev. A **41**, 3847 (1990).

[42] D. Bover, Moment equation methods for nonlinear stochastic systems, J. Math. Anal. Appl. **65**, 306 (1978).

[43] P. J. H. Green, Characteristic functions by e. lukacs. [second edition. Pp. viii + 350. London: Griffin, 1970, £5·50], Journal of the Institute of Actuaries **97**, 134 (1971).

[44] R. Belousov and E. G. D. Cohen, Second-order fluctuation theory and time autocorrelation function for currents, Phys. Rev. E **94**, 062124 (2016).

[45] E. S. Nascimento and W. A. M. Morgado, Energy exchanges in a damped langevin-like system with two thermal baths and an athermal reservoir, J. Phys. A: Math. Theor. **55**, 395003 (2022).

[46] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability (Springer Berlin Heidelberg, 2011).

[47] A. Sarracino and A. Vulpiani, On the fluctuation-dissipation relation in non-equilibrium and non-Hamiltonian systems, Chaos **29**, 083132 (2019).

[48] M. S. Gutiérrez and V. Lucarini, On some aspects of the response to stochastic and deterministic forcings, J. Phys. A: Math. Theor. **55**, 425002 (2022).

[49] D. P. Kingma and M. Welling, Auto-encoding variational bayes, in *Proc. of the 2nd International Conference on Learning Representations (ICLR 2014), Banff, AB, Canada*, edited by Y. Bengio and Y. LeCun (2014).

[50] G. A. Pavliotis and A. Zanoni, Eigenfunction martingale estimators for interacting particle systems and their mean-field limit, SIAM Journal on Applied Dynamical Systems **21**, 2338 (2022).

[51] C. J. Dsilva, R. Talmon, C. W. Gear, R. R. Coifman, and I. G. Kevrekidis, Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems, SIAM Journal on Applied Dynamical Systems **15**, 1327 (2016).

[52] M. Shiino, Dynamical behavior of stochastic systems of infinitely many coupled nonlinear oscillators exhibiting phase transitions of mean-field type: *H* theorem on asymptotic approach to equilibrium and critical slowing down of order-parameter fluctuations, Phys. Rev. A **36**, 2393 (1987).

[53] J. A. Carrillo, K. Craig, and Y. Yao, *Aggregation-Diffusion Equations: Dynamics, Asymptotics, and Singular Limits* (Springer International Publishing, Cham, 2019), pp. 65–108.

[54] J. A. Carrillo, R. S. Gvalani, G. A. Pavliotis, and A. Schlichting, Long-time behaviour and phase transitions for the mckean–vlasov equation on the torus, Arch. Ration. Mech. Anal. **235**, 635 (2020).

[55] F. Malrieu, Logarithmic sobolev inequalities for some nonlinear PDE's, Stochastic Processes and their Applications **95**, 109 (2001).

[56] G. A. Pavliotis, *Stochastic Processes and Applications* (Springer, New York, 2014), Vol. 60.

[57] Y. Klimontovich, Ito, stratonovich and kinetic forms of stochastic equations, Physica A **163**, 515 (1990).

[58] N. G. van Kampen, Itô versus stratonovich, J. Stat. Phys. **24**, 175 (1981).

[59] G. Pesce, A. McDaniel, S. Hottovy, J. Wehr, and G. Volpe, Stratonovich-to-itôtransition in noisy systems with multiplicative feedback, Nat. Commun. **4**, 2733 (2013).