



Response theory and phase transitions for the thermodynamic limit of interacting identical systems

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Lucarini, V. ORCID: <https://orcid.org/0000-0001-9392-1471>, Pavliotis, G. A. and Zagli, N. (2020) Response theory and phase transitions for the thermodynamic limit of interacting identical systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. ISSN 1364-5021 doi: <https://doi.org/10.1098/rspa.2020.0688> Available at <http://centaur.reading.ac.uk/94598/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1098/rspa.2020.0688>

Publisher: Royal Society Publishing

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other

copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Research



Cite this article: Lucarini V, Pavliotis GA, Zagli N. 2020 Response theory and phase transitions for the thermodynamic limit of interacting identical systems. *Proc. R. Soc. A* **476**: 20200688.
<https://doi.org/10.1098/rspa.2020.0688>

Received: 25 August 2020

Accepted: 25 November 2020

Subject Areas:

applied mathematics, complexity, statistical physics

Keywords:

thermodynamic limit, Kramers–Kronig relations, sum rules, Desai–Zwanzig model, Bonilla–Casado–Morilla model, order–disorder transitions

Author for correspondence:

Valerio Lucarini

e-mail: v.lucarini@reading.ac.uk

Response theory and phase transitions for the thermodynamic limit of interacting identical systems

Valerio Lucarini^{1,2}, Grigorios A. Pavliotis³ and Niccolò Zagli^{1,2,3}

¹Department of Mathematics and Statistics, University of Reading, and ²Centre for the Mathematics of Planet Earth, University of Reading, Reading, UK

³Department of Mathematics, Imperial College London, London, UK

VL, 0000-0001-9392-1471

We study the response to perturbations in the thermodynamic limit of a network of coupled identical agents undergoing a stochastic evolution which, in general, describes non-equilibrium conditions. All systems are nudged towards the common centre of mass. We derive Kramers–Kronig relations and sum rules for the linear susceptibilities obtained through mean field Fokker–Planck equations and then propose corrections relevant for the macroscopic case, which incorporates in a self-consistent way the effect of the mutual interaction between the systems. Such an interaction creates a memory effect. We are able to derive conditions determining the occurrence of phase transitions specifically due to system-to-system interactions. Such phase transitions exist in the thermodynamic limit and are associated with the divergence of the linear response but are not accompanied by the divergence in the integrated autocorrelation time for a suitably defined observable. We clarify that such endogenous phase transitions are fundamentally different from other pathologies in the linear response that can be framed in the context of critical transitions. Finally, we show how our results can elucidate the properties of the Desai–Zwanzig model and of the Bonilla–Casado–Morillo model, which feature paradigmatic equilibrium and non-equilibrium phase transitions, respectively.

© 2020 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

1. Introduction

Multi-agent systems are used routinely to model phenomena in the natural sciences, social sciences, and engineering. In addition to the standard applications of interacting particle systems to, e.g. plasma physics and stellar dynamics, phenomena such as cooperation [1], synchronization [2], systemic risk [3], consensus opinion formation [4,5] can be modelled using interacting multi-agent systems. Multi-agent systems are finding applications also in areas like management of natural hazards [6] and of climate change impacts [7]. We refer to [8] for a recent review on interacting multi-agent systems and their applications to the social sciences, and to [9] for a collections of articles showcasing their application in many different areas of science and technology. Additionally, multi-agent systems are also used as the basis for algorithms for sampling and optimization [10].

In this paper, we focus on a particular class of multi-agent systems, namely weakly interacting diffusions, for which the strength of the interaction between the agents is inversely proportional to the number of agents. Under the assumption of exchangeability, i.e. that the particles are identical, it is well known that one can pass to the limit as the number of agents goes to infinity, i.e. the *mean field limit*. In particular, in this limit the evolution of the empirical measure is described by a nonlinear, non-local Fokker–Planck equation, the *McKean–Vlasov Equation* [11,12]. We refer to [13] for a comprehensive review of the McKean–Vlasov equation from a theoretical physics viewpoint. The class of multi-agent models considered in this paper is sufficiently rich to include models for cooperation, systemic risk, synchronization, biophysics, and opinion formation.

An important feature of weakly interacting diffusions is that in the mean field (thermodynamic) limit they can exhibit phase transitions [1,4,14–16]. Phase transitions are characterized in terms of exchange of stability of non-unique stationary states for the McKean–Vlasov equation at the critical temperature/interaction strength.

In the case of equilibrium systems, such stationary states are associated with critical points of a suitably defined energy landscape. For example, for the Kuramoto model of nonlinear oscillators, at the critical noise strength the uniform distribution (on the torus) becomes unstable and stable localized stationary states emerge (phase-locking), leading to synchronization phase transition [17]. A complete theory of phase transitions for the McKean–Vlasov equation on the torus, that includes the Kuramoto model of synchronization, the Hegselmann–Krause model of opinion formation, the Keller–Segel model of chemotaxis. etc., is presented in [18], see also [19]. The effect of (infinitely) many local minima in the energy landscape on the structure of the bifurcation diagram was studied in [20]. Phase transitions for gradient system with local interactions were studied in [21,22]. Synchronization has been extensively discussed in the scientific literature, see [17,23–28].

(a) Linear response theory

One of the main objectives of this paper is to investigate phase transitions for weakly interacting diffusions by looking at the response of the (infinite dimensional) mean field dynamics to weak external perturbations. We associate the nearing of a phase transition with the setting where a very small cause leads to very large effects, or, more technically, to the breakdown of linear response in the system, as described below.

Linear response theory provides a general framework for investigating the properties of physical systems [29]. Well-known applications of linear response theory include solid-state physics and optics [30] as well as plasma physics and stellar dynamics ([31], ch. 5). Furthermore, the range of systems for which linear response theory is relevant is very vast, see e.g. [32–36]. Recently, many new areas of applications of linear response theory are emerging across different disciplinary areas—see, e.g. a recent special issue [37]—and new formulations of the problem are being presented, where the conceptual separation between acting forcing and observed response is blurred [38]. In particular, recent applications of linear response theory include the prediction of climate response to forcings [39–44]. In modern terms, the goal is to define practical ways

to reconstruct the measure supported time-dependent pullback attractor [45] of the climate by studying the response to perturbations of a suitably defined reference climate state [46].

The mathematical theory of linear response for deterministic systems was developed by Ruelle in the context of Axiom A chaotic systems [47,48]. He provided explicit response formulae and showed that, in the case of dissipative systems, the classical fluctuation–dissipation theorem does not hold, and, as a result, natural and forced fluctuations are intimately different [49]. Ruelle’s results have then been re-examined through a more a functional analytic lens by studying the impacts of the perturbations to the dynamics on the transfer operator [50] and then extended to a more abstract mathematical framework [51–53]. The direct implementation of Ruelle’s formulae is extremely challenging, because of the radically different behaviour of the system along the stable and unstable manifold [54], which is related to the insightful *tout court* criticism of linear response theory by Van Kampen [55], so that alternative strategies have been devised [56,57]. Very promising progresses have been recently obtained in the direction of using directly Ruelle’s formulae thanks to adjoint and shadowing methods [58–60].

Linear response theory and fluctuation–dissipation theorems have long been studied in detail for diffusion processes ([61], [62, ch. 7], [63, ch. 9]), and, more recently, rigorous results have been obtained in this direction [64,65]. An interesting link between response theory for deterministic and stochastic systems has been proposed in [66]. The results presented in [64,65] can be applied to the McKean–Vlasov equation *in the absence of phase transitions* to justify rigorously linear response theory and to establish fluctuation–dissipation results. See also [67] for formal calculations. This is not surprising, since it is well known that, in the absence of phase transitions, fluctuations around the mean field limit are Gaussian and can be described in terms of an appropriate stochastic heat equation [1,68].

(b) Critical transitions versus phase transitions

Critical transitions appear when the spectral gap of the transfer operator [51] of the unperturbed system becomes vanishingly small, as a result of the Ruelle–Pollicott poles [69,70] touching the real axis. Since there is a one-to-one correspondence between the radius of expansion of linear response theory and the spectral gap of the transfer operator [51,71], near critical transitions the linear response breaks down and one finds rough dependence of the system properties on its parameters [72,73]. Systems undergoing critical transitions appear often in the natural and social sciences [74] and a lot of effort has been put in the development of early warning signals for critical transitions [75–77]. Early warning signals include an increase in variance and correlation time as the system approaches the transition point.

In the deterministic case, at the critical transition the reference state loses stability and the system ends up in a possibly very different metastable state. Indeed, the presence of critical transitions is closely related to the existence of regimes of multi-stability [78,79]. Transition points for finite-dimensional stochastic systems correspond to points where the topological structure of the *unique* invariant measure changes ([80], [63, Sect. 5.4]). Contrary to this, more than one invariant measure can exist in the mean field (thermodynamic) limit, e.g. in equilibrium systems, when the free energy is not convex [18]. The loss of uniqueness of the stationary state at the critical temperature/noise strength corresponds to a phase transition.

Phase transitions are usually defined by (a) identifying an order parameter and (b) verifying that in the thermodynamic limit, for some value of the parameter of the system, the properties of such an order parameter undergo a sudden change. It should be emphasized, however, that, for the mean field dynamics it is not always possible to identify an order parameter. The way we define phase transitions in this work comes from a somewhat complementary viewpoint, which aims at clarifying analogies and differences with respect to the case of critical transitions.

Sornette and collaborators have devoted efforts at separating the effects of endogeneous versus exogenous processes in determining the dynamics of a complex system and, especially in defining the conditions conducive to crises [81], and proposed multiple applications in the natural (e.g. [82]) as well as the social (e.g. [83]) sciences. The existence of a relationship between the

response of the system to exogenous perturbations and the decorrelation due to endogenous dynamics is interpreted as resulting from a fluctuation–dissipation relation-like properties. Finally, Sornette and collaborators have also emphasized the importance of memory effects especially in the context of endogenous dynamics [84,85]. While our viewpoint and methods are different from theirs, what we pursued here shares similar goals and delves into closely related concepts.

(c) This paper: goals and main results

The main objective of this paper is to perform a systematic study of linear response theory for mean field partial differential equations (PDEs) exhibiting phase transitions. Indeed, it has been shown that, for nonlinear oscillators coupled linearly with their mean, the so-called Desai–Zwanzig model [86], the fluctuations at the phase transition point are not Gaussian [1], see also [19] for related results for a variant of the Kuramoto model (the Haken–Kelso–Bunz model). Indeed, the fluctuations are persistent, non-Gaussian in time, with an amplitude described by a nonlinear stochastic differential equation, and associated with a longer timescale [1]. At the transition point, the standard form of linear response theory breaks down [14]. More general analyses performed using ideas from linear response theory of how a system of coupled maps performs a transition to a coherent state in the thermodynamic limit can be found in [87,88].

Here, we consider a network of N identical and coupled M -dimensional systems whose evolution is described by a Langevin equation. We then study the response to perturbations in the limit of $N \rightarrow \infty$. We investigate the conditions determining the breakdown of the linear response and separate two possible scenarios. One scenario pertains to the closure of the spectral gap of the transfer operator of the mean field equations, and can be dealt with through the classical theory of critical transitions. A second scenario of breakdown of the linear response results from the coupling among the N systems and is inherently associated with the thermodynamic limit. We focus on the second scenario of breakdown of the linear response, which we interpret as corresponding to a phase transition. The main results of this paper can be summarized as follows:

- the derivation of linear response formulae for the thermodynamic limit of a network of coupled identical systems and of Kramers–Kronig relations and sum rules for the related susceptibilities;
- the statement of conditions leading to phase transitions as opposed to the classical scenario of critical transitions;
- the explicit derivation of the corrections to the standard Kramers–Kronig relations and sum rules occurring at the phase transition;
- the clarification, through the use of functional analytical arguments, of why one does not expect divergence of the integrated autocorrelation time of suitable observables in the case of phase transitions, whereas the opposite holds in the case of critical transitions;
- the re-examination, also through numerical simulations, of classical results on phase transitions in the Desai–Zwanzig model [86] and in the Bonilla–Casado–Morillo model [89].

The rest of the paper is organized as follows. In §2, we introduce our model and present the linear response formulae for the mean field equations as well as for the renormalized macroscopic case. In §3, we discuss the properties of the frequency-dependent susceptibility, present the Kramers–Kronig relations connecting their real and imaginary parts, and find explicit sum rules. In §4 we discuss under which conditions the response diverges, and clarify the fundamental difference between the case of critical transitions and the case of phase transitions, which can take place only in the thermodynamic limit. Section 5 is dedicated to finding results that specifically apply to the case of gradient systems, corresponding to reversible Markovian dynamics. In §6 we re-examine the case of phase transitions for the Desai–Zwanzig and Bonilla–Casado–Morilla models, which are relevant for the case of equilibrium and non-equilibrium

dynamics, respectively. Finally, in §7, we present our conclusions and provide perspectives for future investigations.

2. Linear response formulae: mean field and macroscopic results

We consider a network of N exchangeable interacting M -dimensional systems whose dynamics is described by the following stochastic differential equations:

$$d\mathbf{x}_i^k = F_{i,\alpha}(\mathbf{x}^k)dt - \frac{\theta}{N} \sum_{l=1}^N \partial_{\mathbf{x}_i^k} U(\mathbf{x}^k - \mathbf{x}^l) dt + \sigma s_{ij}(\mathbf{x}) dW_j, \quad k=1, \dots, N \quad i=1, \dots, M \quad (2.1)$$

where \mathbf{F}_α is a smooth vector field, possibly depending on a parameter α . Additionally, dW_i , $i=1, \dots, N$ are independent Brownian motions (the Ito convention is used); s_{ij} is the volatility matrix, and the parameter $\sigma > 0$ controls the intensity of the stochastic forcing. Additionally, the N systems undergo an all-to-all coupling through the Laplacian matrix given by the derivative of the potential $U(\mathbf{y}) = |\mathbf{y}|^2/2$. We emphasize the fact that the linear response theory calculations presented below are valid for arbitrary choices of the interaction potential. We choose to present our results for the case of quadratic interactions since in this case the order parameter is known; furthermore, the stationary state(s) are known and are parametrized by the order parameter [1]. The coefficient θ modulates the intensity of such a coupling, which attempts at synchronizing all systems by nudging them to the centre of mass $1/N \sum_{k=1}^N \mathbf{x}^k$. If $\theta = 0$, the N systems are decoupled. We remark that the theory of synchronization says that for this choice of the coupling, if $d\mathbf{x}/dt = \mathbf{F}_\alpha(\mathbf{x})$ has a unique attractor and is chaotic with $\lambda_1 > 0$ being the largest Lyapunov exponent, the N nodes undergo perfect synchronization for any $N \geq 2$ in the absence of noise ($\sigma = 0$) if $\theta > \lambda_1$ [17,28,90,91].

If $\mathbf{F}_\alpha(\mathbf{y}) = -\nabla V_\alpha(\mathbf{y})$, we interpret V_α as the confining potential [63]. In some cases, equation (2.1) describes an equilibrium statistical mechanical system, in particular if $\mathbf{F}_\alpha = -\nabla V_\alpha(\mathbf{y})$ and s_{ij} is proportional to the identity. More generally, equilibrium conditions are realized when the drift term—the deterministic component on the right-hand side of equation (2.1)—is proportional to the gradient of a function defined according to the Riemannian metric given by the diffusion matrix $C_{ij} = s_{ik}s_{jk}$ [92].

We now consider the empirical measure $\rho^{(N)}$, which is defined as $\rho^{(N)} = 1/N \sum_{k=1}^N \delta_{\mathbf{x}^k(t)}$. Following [11,18,93], we investigate the thermodynamic limit of the system above. As $N \rightarrow \infty$, we can use martingale techniques [1,11,93,94] to show that the one-particle density converges to some measure $\rho(\mathbf{x}, t)$ satisfying the following McKean–Vlasov equation, which is a nonlinear and non-local Fokker–Planck equation :

$$\begin{aligned} \frac{\partial \rho(\mathbf{x}, t)}{\partial t} &= -\nabla \cdot [\rho(\mathbf{x}, t) (\mathbf{F}_\alpha(\mathbf{x}) - \theta \nabla U \star \rho)] + \frac{\sigma^2}{2} \tilde{\Delta} \rho(\mathbf{x}, t) \\ &= -\nabla \cdot [\rho(\mathbf{x}, t) (\mathbf{F}_\alpha(\mathbf{x}) + \theta (\langle \mathbf{x} \rangle(t) - \mathbf{x}))] + \frac{\sigma^2}{2} \tilde{\Delta} \rho(\mathbf{x}, t) \\ &:= L_{\alpha,\theta}^0(\rho(\mathbf{x}, t)) + \theta \Lambda_\theta(\{\rho(\mathbf{x}, t)\}), \end{aligned} \quad (2.2)$$

where we have separated the linear operator $L_{\alpha,\theta}^0$ and the nonlinear operator $\Lambda_\theta(\{\rho(\mathbf{x}, t)\}) = \theta \nabla \cdot (\rho(\mathbf{x}, t) \langle \mathbf{x} \rangle(t))$, with $\langle \mathbf{x} \rangle(t) = \int d^M \mathbf{y} \rho(\mathbf{y}, t)$ and \star denotes the convolution. Additionally, we have that $\tilde{\Delta}$ is a linear diffusion operator such that $\tilde{\Delta} \rho(\mathbf{x}, t) = \sum_{i=1}^M \sum_{j=1}^M \partial_{x_i} \partial_{x_j} (C_{ij}(\mathbf{x}) \rho(\mathbf{x}, t))$, which coincides with the standard M -dimensional Laplacian ($\tilde{\Delta} = \Delta$) if the diffusion matrix C_{ij} is the identity matrix. If $\sigma = 0$, we are considering a nonlinear Liouville equation. We assume that, if $\sigma > 0$, equation (2.1) describes a hypoelliptic diffusion process, so that $\rho_{\alpha,\theta}^{(0)}(\mathbf{x})$ is smooth [63, ch. 6]. In what follows, we refer to the case $\sigma > 0$. Conditions detailing the well-posedness of this problem can be found in [95].

Let us define $\rho_{\alpha,\beta}^{(0)}(\mathbf{x})$ as a reference invariant measure of the system such that $L_{\alpha,\beta}^0(\rho_{\alpha,\beta}^{(0)}(\mathbf{x})) + \theta \Lambda_\theta(\{\rho_{\alpha,\beta}^{(0)}(\mathbf{x})\}) = 0$. Since we are considering a system with an infinite number of particles, such an invariant measure needs not be unique [1,18,96,97]. Specifically, if s_{ij} is proportional to the identity and $\mathbf{F}_\alpha(\mathbf{y}) = -\nabla V_\alpha(\mathbf{y})$ and $V_\alpha(\mathbf{y})$ is not convex, thus allowing for more than one local minimum, for a given value of θ the system undergoes a phase transition for sufficiently weak noise; see discussion in §5.

We remark that the invariant measure depends on the values of α and θ , and, in particular, $\langle \mathbf{x} \rangle(t) = \langle \mathbf{x} \rangle_{\alpha,\beta}^{(0)} = \langle \mathbf{x} \rangle_0$ is a constant vector, where in the last identity we have dropped the lower indices to simplify the notation. As a result, we have that:

$$M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0(\rho_{\alpha,\beta}^{(0)}(\mathbf{x})) = -\nabla \cdot (\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) (\mathbf{F}(\mathbf{x}) + \theta (\langle \mathbf{x} \rangle_0 - \mathbf{x}))) + \frac{\sigma^2}{2} \tilde{\Delta} \rho_{\alpha,\beta}^{(0)}(\mathbf{x}) = 0 \quad (2.3)$$

so that the invariant measure $\rho_{\alpha,\beta}^{(0)}(\mathbf{x})$ is the eigenvector with vanishing eigenvalue of the linear operator $M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0$.

Taking inspiration from [98,99], we now study the impact of perturbations on the invariant measure $\rho_{\alpha,\beta}^{(0)}(\mathbf{x})$. We follow and extend the results presented in [67]. We modify the right-hand side of equation (2.2) by setting $\mathbf{F}_\alpha(\mathbf{x}) \rightarrow \mathbf{F}_\alpha(\mathbf{x}) + \epsilon \mathbf{X}(\mathbf{x})T(t)$ and we study the linear response of the system in terms of the density $\rho(\mathbf{x}, t)$. We then write $\rho(\mathbf{x}, t) = \rho_{\alpha,\beta}^{(0)}(\mathbf{x}) + \epsilon \rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t) + o(\epsilon^2)$ and obtain the following equation up to order ϵ :

$$\begin{aligned} \frac{\partial \rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t)}{\partial t} &= M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0(\rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t)) - T(t) \nabla \cdot (\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) \mathbf{X}(\mathbf{x})) \\ &\quad - \theta \nabla \cdot (\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(1)}(\mathbf{y}, t) \mathbf{y}) \\ &= \tilde{M}_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0(\rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t)) - T(t) \nabla \cdot (\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) \mathbf{X}(\mathbf{x})) \end{aligned} \quad (2.4)$$

We remark that the linear operator $\tilde{M}_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0$ acting on $\rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t)$ on the right-hand side of the previous equation is not the operator whose zero eigenvector is the unperturbed invariant measure. The correction proportional to θ emerges as a result of the nonlinearity of the McKean–Vlasov equation. We will discuss the operator $\tilde{M}_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0$ in §2a.

One then derives:

$$\begin{aligned} \rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t) &= \int_{-\infty}^t ds \exp(M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0(t-s)) \left[-T(s) \nabla \cdot (\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) \mathbf{X}(\mathbf{x})) \right] \\ &\quad + \int_{-\infty}^t ds \exp(M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^0(t-s)) \left[-\theta \nabla \cdot (\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(1)}(\mathbf{y}, s) \mathbf{y}) \right] \end{aligned} \quad (2.5)$$

We now evaluate the response of the observable x_i . This is sufficient for our purposes, since we know that, for this model, the order parameter (in the mean field limit) is the mean position $\langle \mathbf{x} \rangle$ (magnetization). By definition, we have that

$$\langle x_i \rangle(t) = \langle x_i \rangle_0 + \langle x_i \rangle_1(t) + O(\epsilon^2),$$

where we have defined $\langle \Phi \rangle_0 = \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(0)}(\mathbf{y}) \Phi(\mathbf{y})$ and $\langle \Phi \rangle_1(t) = \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(1)}(\mathbf{y}, t) \Phi(\mathbf{y})$ for a generic observable Φ . We obtain:

$$\begin{aligned} \langle x_i \rangle_1(t) &= \int_{-\infty}^t ds \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y}) T(s) \cdot \nabla \exp(M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^{0,+}(t-s)) y_i \\ &\quad + \theta \int_{-\infty}^t ds \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(0)}(\mathbf{y}) \langle \mathbf{x} \rangle_1(s) \cdot \nabla \exp(M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^{0,+}(t-s)) y_i, \end{aligned} \quad (2.6)$$

where we have defined the following operator:

$$M_{\alpha,\beta,\langle \mathbf{x} \rangle_0}^{0,+} = \mathbf{F}(\mathbf{x}) \cdot \nabla + \theta (\langle \mathbf{x} \rangle_0 - \mathbf{x}) \cdot \nabla + \frac{\sigma^2}{2} \tilde{\Delta}^+ \quad (2.7)$$

where O^+ is the adjoint of O . Following [67], we can interpret this as the Koopman operator for the unperturbed dynamics; see later discussion. We can rewrite the previous expression as:

$$\langle x_i \rangle_1(t) = \int_{-\infty}^{\infty} ds T(s) G_{i,\alpha,\theta}(t-s) + \sum_{k=1}^M \int_{-\infty}^{\infty} ds \langle x_k \rangle_1(s) Y_{\{i,k\},\alpha,\theta}(t-s) \quad (2.8)$$

where

$$G_{i,\alpha,\theta}(\tau) = \Theta(\tau) \int d^M \mathbf{y} \left(\rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y}) \right) \cdot \nabla \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}(\tau) \right) y_i \quad (2.9)$$

and

$$Y_{\{i,k\},\alpha,\theta}(\tau) = \theta \Theta(\tau) \int d^M \mathbf{y} \rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \partial_{y_k} \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}(\tau) \right) y_i, \quad (2.10)$$

where the Green function is causal. Note also that if $\mathbf{X}(\mathbf{x}) = \hat{\mathbf{v}}_k$, where $\hat{\mathbf{v}}_k$ is the unit vector in the k th direction, then $G_{i,\alpha,\theta}(\tau) = Y_{\{i,k\},\alpha,\theta}(\tau)/\theta$.

Notwithstanding the Markovianity of the dynamics, the second term on the right-hand side of equation (2.8) describes a memory effect in the response of the observable x . Such a term emerges in the thermodynamic limit and effectively imposes a condition of self-consistency between forcing and response; see different yet related results obtained by Sornette and collaborators [81,84,85].

If $\sigma > 0$ the invariant measure is smooth, so that we can perform an integration by parts of the previous expressions and derive the following Green functions:

$$G_{i,\alpha,\theta}(\tau) = -\Theta(\tau) \int d^M \mathbf{y} \rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \frac{\nabla \cdot \left(\rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y}) \right)}{\rho_{\alpha,\theta}^{(0)}(\mathbf{y})} \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}(\tau) \right) y_i \quad (2.11)$$

and

$$Y_{\{i,k\},\alpha,\theta}(\tau) = -\theta \Theta(\tau) \int d^M \mathbf{y} \rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \partial_{y_k} \log \left(\rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \right) \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}(\tau) \right) y_i, \quad (2.12)$$

where the Green functions are written as correlation functions times a Heaviside distribution enforcing causality.

We remark that we can, at least formally, write:

$$\exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}(t) \right) = \Pi_0 + \sum_{j=1}^{\infty} \exp(t\lambda_j) \Pi_j + \mathcal{R}(t), \quad (2.13)$$

where $\{\lambda_j\}_{j=1}^{\infty}$ are the eigenvalues (point-spectrum) of $M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}$ and Π_j is the spectral projector onto the eigenspace spanned by the eigenfunction ψ_j , and in particular, Π_0 projects on the invariant measure. Then, the operator $\mathcal{R}(t)$ is the residual operator associated with the essential spectrum. The norm of $\mathcal{R}(t)$ is controlled by the distance of essential spectrum from the imaginary axis.

We then have:

$$G_{i,\alpha,\theta}(\tau) = \Theta(\tau) \sum_{j=1}^{\infty} \langle \psi_j y_i \rangle_0 \langle \Phi_X \psi_j \rangle_0 \exp(\lambda_j t) + \mathcal{R}_{\Phi_X}(\tau) \quad (2.14)$$

and

$$Y_{\{i,k\},\alpha,\theta}(\tau) = \Theta(\tau) \sum_{j=1}^{\infty} \langle \psi_j y_i \rangle_0 \langle \Phi_k \psi_j \rangle_0 \exp(\lambda_j t) + \mathcal{R}_{\Phi_k}(\tau), \quad (2.15)$$

where $\Phi_X = -\nabla \cdot \left(\rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y}) \right) / \rho_{\alpha,\theta}^{(0)}(\mathbf{y})$ and $\Phi_k = -\theta \partial_{y_k} \log(\rho_{\alpha,\theta}^{(0)}(\mathbf{y}))$. Note that the $j=0$ term vanishes because the corresponding scalar product $\langle \Phi_X \psi_0 \rangle_0$ has nil value for any choice of the vector field \mathbf{X} .

We now apply the Fourier transform \mathcal{F} to equation (2.8) and obtain:

$$P_{ij,\alpha,\theta}(\omega)\langle x_j \rangle_1(\omega) = \Gamma_{i,\alpha,\theta}(\omega)T(\omega) \quad P_{ij,\alpha,\theta}(\omega) = \delta_{ij} - \Upsilon_{\{i,j\},\alpha,\theta}(\omega) \quad (2.16)$$

where we have used a (standard) abuse of notation in defining the Fourier transform of $T(t)$ and $\langle x_j \rangle_1(t)$ and have defined

$$\Gamma_{i,\alpha,\theta}(\omega) = \mathcal{F}\{G_{i,\alpha,\theta}(\omega)\} = \sum_{j=1}^{\infty} \frac{\langle \psi_j y_i \rangle_0 \langle \Phi_X \psi_j \rangle_0}{i\omega + \lambda_j} + \mathcal{R}_{\Phi_X}(\omega) \quad (2.17)$$

and

$$\Upsilon_{\{i,k\},\alpha,\theta}(\omega) = \mathcal{F}\{Y_{\{i,k\},\alpha,\theta}(t)\} = \sum_{j=1}^{\infty} \frac{\langle \psi_j y_i \rangle_0 \langle \Phi_k \psi_j \rangle_0}{i\omega + \lambda_j} + \mathcal{R}_{\Phi_k}(\omega). \quad (2.18)$$

We remark that the susceptibilities given in equations (2.17) and (2.18) are holomorphic in the upper complex ω -plane if $\mathbf{Re}\{\lambda_j\} < 0$, $j = 1, \dots, \infty$. Note that all susceptibilities, regardless of the observable considered, share the same poles located at $\omega_j = i\lambda_j$, $j = 1, \dots, \infty$. Additionally, if ω_j is a pole, so is also $-\omega_j^*$ (and, correspondingly, λ_j comes together with λ_j^*).

By introducing the inverse matrix $\Pi_{\alpha,\theta} = P_{\alpha,\theta}^{-1}$, we obtain from equation (2.16) our final result:

$$\langle x_i \rangle_1(\omega) = \Pi_{ij,\alpha,\theta}(\omega)\Gamma_{i,\alpha,\theta}(\omega)T(\omega) = \tilde{\Gamma}_{i,\alpha,\theta}(\omega)T(\omega) \quad (2.19)$$

where:

$$\tilde{\Gamma}_{i,\alpha,\theta}(\omega) = \Pi_{ij,\alpha,\theta}(\omega)\Gamma_{j,\alpha,\theta}(\omega). \quad (2.20)$$

The previous expression generalizes previous findings presented in [87]. We will discuss below the invertibility properties of the matrix $P_{ij,\alpha,\theta}(\omega)$. If the coupling is absent, so that $\theta = 0$, we obtain the same result as in the case of a single particle $N = 1$ system: $\langle x_i \rangle_1(\omega) = \Gamma_{i,\alpha,\theta=0}(\omega)T(\omega)$. Additionally, we trivially get $\Gamma_{i,\alpha,\theta=0}(\omega) = \tilde{\Gamma}_{i,\alpha,\theta=0}(\omega)$. The effect of switching on the coupling and taking $\theta > 0$ is twofold in terms of response:

- First, the function $\Gamma_{i,\alpha,\theta}(\omega)$ is modified, because the unperturbed evolution operator $M_{\alpha,\theta,(x)_0,\theta}^0$ (see equation (2.4)) and the unperturbed invariant measure $\rho_{\alpha,\theta}^{(0)}(\mathbf{x})$ depend explicitly on θ . Indeed, changes in the value of θ impact expectation values and correlation properties. From the definition of $M_{\alpha,\theta,(x)_0,\theta}^0$, we interpret $\Gamma_{i,\alpha,\theta}(\omega)$ as the mean field susceptibility.
- More importantly, the presence of a non-vanishing value of θ introduces a non-trivial correction with respect to the identity to the matrix $P_{ij,\alpha,\theta}(\omega)$. We can interpret the function $\tilde{\Gamma}_{i,\alpha,\theta}(\omega)$ as the macroscopic susceptibility, which takes fully into account, in a self-consistent way, the interaction between the systems. Equation (2.19) generalizes the frequency-dependent version of the well-known Clausius–Mossotti relation [30,100, 101], which connects the macroscopic polarizability of a material and the microscopic polarizability of its elementary components.

The integration by parts used for deriving equations (2.11) and (2.12) from equations (2.9) and (2.10) amounts to deriving a variant of the fluctuation–dissipation relation [29,33], as the Green functions are written as the causal part of a time-lagged correlation of two observables as determined by unperturbed dynamics. In other terms, the poles ω_j , $j = 1, \dots, \infty$ of the susceptibilities above correspond to the Ruelle–Pollicott poles [69,70] of the unperturbed system, just as in the case of systems described by the standard Fokker–Planck equation [73,102]. This establishes a close connection between forced and free variability or, using a different terminology, between the properties of response to exogenous perturbations and endogenous dynamics [81].

(a) Another expression for the macroscopic susceptibility

A somewhat unsatisfactory aspect of the previous derivation resides in the fact that we are dealing with the operator $\exp(M_{\alpha,\beta,(x)_0}^{0,+} t)$, which is associated with the mean field approximation. We can instead proceed from equation (2.5) using the operator $\exp(\tilde{M}_{\alpha,\beta,(x)_0}^0 t)$ introduced above and derive directly the following results:

$$\rho_{\alpha,\beta}^{(1)}(\mathbf{x}, t) = \int_{-\infty}^t ds \exp\left(\tilde{M}_{\alpha,\beta,(x)_0}^0(t-s)\right) \left[-T(s) \nabla \cdot \left(\rho_{\alpha,\beta}^{(0)}(\mathbf{x}) \mathbf{X}(\mathbf{x})\right)\right] \quad (2.21)$$

and

$$\langle x_i \rangle_1(t) = \int_{-\infty}^t ds \int d^M \mathbf{y} \rho_{\alpha,\beta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y}) T(s) \cdot \nabla \exp\left(\tilde{M}_{\alpha,\beta,(x)_0}^{0,+}(t-s)\right) y_i T(s). \quad (2.22)$$

We can rewrite the previous expression as:

$$\langle x_i \rangle_1(t) = \int_{-\infty}^{\infty} ds T(s) \tilde{G}_{i,\alpha,\beta}(t-s), \quad (2.23)$$

where the Fourier transform of:

$$\begin{aligned} \tilde{G}_{i,\alpha,\beta}(\tau) &= \Theta(\tau) \int d^M \mathbf{y} \left(\rho_{\alpha,\beta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y})\right) \cdot \nabla \exp\left(\tilde{M}_{\alpha,\beta,(x)_0}^{0,+}(\tau)\right) y_i \\ &= -\Theta(\tau) \int d^M \mathbf{y} \nabla \cdot \left(\rho_{\alpha,\beta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y})\right) \exp\left(\tilde{M}_{\alpha,\beta,(x)_0}^{0,+}(\tau)\right) y_i \end{aligned} \quad (2.24)$$

is the macroscopic susceptibility introduced in equation (2.19). Note that $\tilde{M}_{\alpha,\beta,(x)_0}^{0,+}$ cannot be interpreted as the generator of time translation for smooth observables.

Clearly, the benefit of deriving the expression of $\tilde{\Gamma}_{i,\alpha,\beta}(\omega)$ as done in the previous section lies in the possibility of bypassing the space-integral operator included in the definition of $\tilde{M}_{\alpha,\beta,(x)_0}^{0,+}$. Similar to equation (2.13), we can write:

$$\exp\left(\tilde{M}_{\alpha,\beta,(x)_0}^{0,+} t\right) = \sum_{j=1}^{\infty} \exp\left(t \tilde{\lambda}_j\right) \tilde{\Pi}_j + \tilde{\mathcal{R}}(t), \quad (2.25)$$

where the corresponding symbols are used. We then have:

$$\tilde{\Gamma}_{i,\alpha,\beta}(\tau) = \Theta(\tau) \sum_{j=1}^{\infty} \langle \tilde{\psi}_j y_i \rangle_0 \langle \Phi_{\Gamma} \tilde{\psi}_j \rangle_0 \exp\left(\tilde{\lambda}_j \tau\right) + \tilde{\mathcal{R}}_{\Phi_X}(\tau). \quad (2.26)$$

We now apply the Fourier transform to equation (2.26) and obtain:

$$\tilde{\Gamma}_{i,\alpha,\beta}(\omega) = \sum_{j=1}^{\infty} \frac{\langle \psi_j y_i \rangle_0 \langle \Phi_X \psi_j \rangle_0}{i\omega + \tilde{\lambda}_j} + \tilde{\mathcal{R}}_{\Phi_{\Gamma}}(\omega) \quad (2.27)$$

Comparing equations (2.27) and (2.20), it is clear that the poles $\tilde{\omega}_j$ of $\tilde{\Gamma}_{i,\alpha,\beta}(\omega)$ are those of $\Gamma_{i,\alpha,\beta}(\omega)$ plus those of the matrix $\Pi_{ij,\alpha,\beta}(\omega)$, see earlier comments by Dawson [1] for the case of the Desai-Zwanzig model [86] (see also §6a).

3. Dispersion relations far from criticalities

We assume that all $\lambda_j, j = 1, \dots, \infty$ have negative real part. As discussed above, since $G_{j,\alpha,\beta}^{(1)}(\tau)$ is causal, the function $\Gamma_{j,\alpha,\beta}^{(1)}(\omega)$ is a well-behaved susceptibility function that is holomorphic in the upper complex ω -plane ($\text{Im}\{\omega\} \geq 0$).

Let us now consider the short-time behaviour $\tau \rightarrow 0^+$ of the response functions $G_{i,\alpha,\beta}(\tau)$. Using equations (2.7) and (2.9), we derive:

$$G_{i,\alpha,\beta}(\tau) = \Theta(\tau) \left(\langle X_i(\mathbf{x}) \rangle_0 + \left(\sum_{k=1}^M \langle X_k(\mathbf{x}) \partial_{x_k} F_i(\mathbf{x}) \rangle_0 - \theta \langle X_i(\mathbf{x}) \rangle_0 \right) \tau + o(\tau^2) \right). \quad (3.1)$$

As a result, the high-frequency behaviour of the susceptibility $\Gamma_{i,\alpha,\beta}(\omega)$ can be written as:

$$\Gamma_{i,\alpha,\beta}(\omega) = i \frac{\langle X_i(\mathbf{x}) \rangle_0}{\omega} - \frac{\sum_{k=1}^M \langle X_k(\mathbf{x}) \partial_{x_k} F_i(\mathbf{x}) \rangle_0 - \theta \langle X_i(\mathbf{x}) \rangle_0}{\omega^2} + o(\omega^2). \quad (3.2)$$

The causality of $G_{i,\alpha,\beta}(\tau)$ implies that, using an abuse of notation, $G_{i,\alpha,\beta}(\tau) = \Theta(\tau)G_{i,\alpha,\beta}(\tau)$. By performing the Fourier transform of both sides of this identity, we obtain the following identity $\Gamma_{i,\alpha,\beta}(\omega) = \frac{1}{2\pi} \Gamma_{i,\alpha,\beta}(\omega) \star \tilde{\Theta}(\omega)$, where \star indicates the convolution product and $\tilde{\Theta}(\omega) = -i\mathbf{P}(1/\omega) + \pi\delta(\omega)$ is the Fourier transform of $\Theta(\tau)$, with \mathbf{P} indicating the principal part. By separating the real (**Re**) and imaginary (**Im**) parts of $\Gamma_{i,\alpha,\beta}(\omega)$, the previous relation can be written as:

$$\mathbf{P} \int_{-\infty}^{\infty} dv \frac{\mathbf{Re}\{\Gamma_{i,\alpha,\beta}(v)\}}{v - \omega} = -\pi \mathbf{Im}\{\Gamma_{i,\alpha,\beta}(\omega)\} \quad (3.3)$$

and

$$\mathbf{P} \int_{-\infty}^{\infty} dv \frac{\mathbf{Im}\{\Gamma_{i,\alpha,\beta}(v)\}}{v - \omega} = \pi \mathbf{Re}\{\Gamma_{i,\alpha,\beta}(\omega)\}. \quad (3.4)$$

Since $\Gamma_{i,\alpha,\beta}(\tau)$ is a real function of real argument τ , its Fourier transform obeys the following conditions: $\Gamma_{i,\alpha,\beta}(\omega) = (\Gamma_{i,\alpha,\beta}(-\omega^*))^*$. Hence, for real values of ω we have $\mathbf{Re}\{\Gamma_{i,\alpha,\beta}(\omega)\} = \mathbf{Re}\{\Gamma_{i,\alpha,\beta}(-\omega)\}$ and $\mathbf{Im}\{\Gamma_{i,\alpha,\beta}(\omega)\} = -\mathbf{Im}\{\Gamma_{i,\alpha,\beta}(-\omega)\}$. We derive an alternative form of the Kramers-Kronig relations [30]:

$$\mathbf{P} \int_0^{\infty} dv \frac{\mathbf{Re}\{\Gamma_{i,\alpha,\beta}(v)\}}{v^2 - \omega^2} = -\frac{\pi}{2\omega} \mathbf{Im}\{\Gamma_{i,\alpha,\beta}(\omega)\} \quad (3.5)$$

and

$$\mathbf{P} \int_0^{\infty} dv \frac{v \mathbf{Im}\{\Gamma_{i,\alpha,\beta}(v)\}}{v^2 - \omega^2} = \frac{\pi}{2} \mathbf{Re}\{\Gamma_{i,\alpha,\beta}(\omega)\}. \quad (3.6)$$

It is then possible to derive the following sum rules:

$$\int_0^{\infty} dv \mathbf{Re}\{\Gamma_{i,\alpha,\beta}(v)\} = \lim_{\omega \rightarrow \infty} \left(\frac{\pi}{2} \omega \mathbf{Im}\{\Gamma_{i,\alpha,\beta}(\omega)\} \right) = \frac{\pi}{2} \langle X_i(\mathbf{x}) \rangle_0 \quad (3.7)$$

and

$$\int_0^{\infty} dv \frac{\mathbf{Im}\{\Gamma_{i,\alpha,\beta}(v)\}}{v} = \lim_{\omega \rightarrow 0} \left(\frac{\pi}{2} \mathbf{Re}\{\Gamma_{i,\alpha,\beta}(\omega)\} \right) = \frac{\pi}{2} \tau_{G_i} G_{i,\alpha,\beta}(0^+), \quad (3.8)$$

where $\tau_{G_i} = \int_0^{\infty} dt G_{i,\alpha,\beta}(t)/G_{i,\alpha,\beta}(0^+)$, if $G_{i,\alpha,\beta}(0^+) \neq 0$ is a measure of the decorrelation of the system, see a related result in [103] on the Desai-Zwanzig model [86] discussed below. Note that $\mathbf{Im}\{\Gamma_{i,\alpha,\beta}(\omega)\}$ is an odd function of ω . Additionally, if $\langle X_i(\mathbf{x}) \rangle_0 = 0$, so that the imaginary part of the susceptibility decreases asymptotically at least as fast as ω^{-3} , the following additional sum rules holds:

$$\int_0^{\infty} dv v \mathbf{Im}\{\Gamma_{i,\alpha,\beta}(v)\} = \lim_{\omega \rightarrow \infty} \left(-\frac{\pi}{2} \omega^2 \mathbf{Re}\{\Gamma_{i,\alpha,\beta}(\omega)\} \right) = \frac{\pi}{2} \sum_{k=1}^M \langle X_k(\mathbf{x}) \partial_{x_k} F_i(\mathbf{x}) \rangle_0. \quad (3.9)$$

Let us now look at the asymptotic properties for large values of ω of the matrix $P_{ij,\alpha,\beta}(\omega)$. We proceed as above and consider the short time behaviour of $Y_{\{ij\},\alpha,\beta}(\tau)$:

$$Y_{\{ij\},\alpha,\beta}(\tau) = \Theta(\tau) (\delta_{ij}\theta + o(\tau)). \quad (3.10)$$

As a result, for large values of ω , we have that

$$\Upsilon_{\{i,k\},\alpha,\beta}(\omega) = i \frac{\theta}{\omega} \delta_{i,k} + o(\omega^{-1}) \quad (3.11)$$

so that $P_{ij,\alpha,\theta}(\omega) = \delta_{ij}(1 - i(\theta/\omega)) + o(\omega^{-2})$ and $\Pi_{ij,\alpha,\theta}(\omega) = \delta_{ij}(1 + i(\theta/\omega)) + o(\omega^{-2})$, so that:

$$\tilde{\Gamma}_{i,\alpha,\theta}(\omega) = i \frac{\langle X_i(\mathbf{x}) \rangle_0}{\omega} - \frac{\sum_{k=1}^M \langle X_k(\mathbf{x}) \partial_{x_k} F_i(\mathbf{x}) \rangle_0}{\omega^2} + o(\omega^2) \quad (3.12)$$

where we note a correction in the asymptotic behaviour with respect to the case of the mean field susceptibility given in equation (3.2). Nonetheless, if $P_{ij,\alpha,\theta}$ has full rank for all values of ω in the upper complex ω -plane, the Kramers–Kronig relations (3.5) and (3.6) and the sum rules (3.7)–(3.9) apply also for the macroscopic susceptibilities $\tilde{\Gamma}_{i,\alpha,\theta}(\omega)$.

4. Criticalities

We remark again that the dispersion relations presented above apply for the mean field susceptibilities for values of α and θ such that (i) the real part of all the eigenvalues of $M_{\alpha,\theta,(\mathbf{x})_0}^{0,+}$ is negative; and for the macroscopic susceptibility if, additionally, (ii) the matrix $P_{ij,\alpha,\theta}$ is invertible and, additionally, has no zeros in the upper complex ω -plane. Conditions (i) and (ii) correspond to the case where the real part of all the eigenvalues of $\tilde{M}_{\alpha,\theta,(\mathbf{x})_0}^{0,+}$ is negative.

The breakdown of condition (i) for, say, $(\alpha, \theta) = (\tilde{\alpha}, \tilde{\theta})$ is due to the presence of a vanishing spectral gap for the operator $M_{\tilde{\alpha},\tilde{\theta},(\mathbf{x})_0}^{0,+}$ and, *a fortiori*, for the operator $\tilde{M}_{\tilde{\alpha},\tilde{\theta},(\mathbf{x})_0}^{0,+}$. In such a scenario, the functions $\Gamma_{i,\tilde{\alpha},\tilde{\theta}}(\omega)$ and $\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\theta}}(\omega)$ feature one or more poles in the real ω -axis. In other terms, linear response blows up for forcings having non-vanishing spectral power $|T(\omega)|^2$ at the corresponding frequencies.

In this case, because of the link discussed above between the poles of the mean field susceptibilities and the Ruelle–Pollicott poles of the unperturbed system, the blow-up of the linear susceptibilities corresponds to an ultraslow decay of correlations leading to a singularity in the integrated decorrelation time. In other terms, in this case the results conform to the classic framework of the theory of critical transitions [46,72,73,76,104]. We remark that the presence of a divergence does not depend on the specific functional form of the perturbation field \mathbf{X} , while the properties of the response do depend in general from it.

The breakdown of condition (ii) for, say, $(\alpha, \theta) = (\tilde{\alpha}, \tilde{\theta})$ is associated with the fact that the spectral gap of the operator $\tilde{M}_{\tilde{\alpha},\tilde{\theta},(\mathbf{x})_0}^{0,+}$ vanishes, while the spectral gap of the operator $M_{\tilde{\alpha},\tilde{\theta},(\mathbf{x})_0}^{0,+}$ remains finite. In this latter case, only the functions $\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\theta}}(\omega)$ have one or more poles for real values of ω , whereas the functions $\Gamma_{i,\tilde{\alpha},\tilde{\theta}}(\omega)$ are holomorphic in the upper complex ω -plane. We remark that the non-invertibility of the P matrix depends on the presence of sufficiently strong coupling between the systems, which leads to them being coordinated, as discussed in detail in §6.

The nonlinearity of equation (2.2) emerges as a result of the thermodynamic limit $N \rightarrow \infty$. Therefore, we interpret the singularities in the linear response resulting from the breakdown of condition (ii) as being associated to a phase transition of the system, yet not a standard one. Indeed, the blow-up of the linear susceptibilities *does not* correspond to a blow-up of the integrated correlation time (see §6a).

(a) Phase transitions

In what follows, we focus on the criticalities associated with condition (ii) only, which emerge *specifically* from effects that cannot be described using the mean field approximation.

Let's then assume that for some reference values for $\alpha = \alpha_0$ and $\theta = \theta_0$ the system is stable. This corresponds to the fact that the inverse Fourier transform of $\tilde{\Gamma}_{i,\alpha_0,\theta_0}(\omega)$, which defines a renormalized linear Green function that takes into account all the interactions among the identical systems, has only positive support. Correspondingly, the macroscopic susceptibilities $\tilde{\Gamma}_{i,\alpha_0,\theta_0}(\omega)$, just like the mean field ones, are holomorphic in the upper complex ω -plane. This implies that the entries of the matrix $\Pi_{ij,\alpha_0,\theta_0}(\omega)$ do not have poles in the upper complex ω -plane.

Let us now consider the following modulation of the system. We consider the protocol $(\alpha_s, \theta_s) = (\alpha_0 + \delta_\alpha(s), \theta_0 + \delta_\theta(s))$ and assume for $0 \leq s < \tilde{s}$ the system retains stability. For $(\alpha_{\tilde{s}}, \theta_{\tilde{s}}) = (\tilde{\alpha}, \tilde{\theta})$,

the system loses stability as R poles ω_l , $l = 1, \dots, R$ cross into the upper complex ω -plane (with $\mathbf{Im}\{\omega_l\} = 0$, $l = 1, \dots, R$) for the macroscopic susceptibilities $\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(\omega)$ (condition ii) is broken), while the mean field susceptibilities $\Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega)$ are holomorphic in the upper complex ω -plane (condition i) holds). This implies that the spectral gap of the operator $M_{\alpha,\theta,(x)_0}^{0,+}$ is finite, so that there is no divergence of the integrated autocorrelation time of any observable.

We have that $P_{ij,\tilde{\alpha},\tilde{\beta}}$ does *not* have full rank for $\omega = \omega_l$, $l = 1, \dots, R$. For such value(s) of ω , the macroscopic susceptibilities diverge. Indeed, we remark that the invertibility conditions of the matrix $P_{ij,\alpha,\beta}(\omega)$ is intrinsic and does not depend on the applied external forcing \mathbf{X} , which enters, instead, only in the definition of the mean field susceptibility $\Gamma_{i,\alpha,\beta}(\omega)$. We interpret this as the fact that the divergence of the response is due to eminently endogenous, rather than exogeneous, processes.

We also remark that $P_{ij,\alpha,\beta}(\omega) = \delta_{ij} - \Upsilon_{[i,j],\alpha,\beta}(\omega)$, where $\Upsilon_{[i,j],\alpha,\beta}(\omega)$ can be seen as mean field susceptibility for the expectation value of x_i associated with an infinitesimal change of the value of the j^{th} component of $\langle \mathbf{x} \rangle_0$, see equations (2.4) and (2.10). This supports the idea that $\langle \mathbf{x} \rangle$ is an appropriate order parameter for the system.

We assume, for simplicity, that only simple poles are present. We then decompose the matrix $\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega)$ in the upper complex ω -plane as follows:

$$\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega) = \Pi_{ij,\tilde{\alpha},\tilde{\beta}}^h(\omega) + \sum_{l=1}^R \frac{\text{Res}(\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega))_{\omega=\omega_l}}{\omega - \omega_l} \quad (4.1)$$

where we have separated the holomorphic component $\Pi_{ij,\tilde{\alpha},\tilde{\beta}}^h(\omega)$ from the singular contributions coming from the poles ω_l , $l = 1, \dots, R$; note that $\text{Res}(f(\omega))_{\omega=v}$ indicates the residue of the function f for $\omega = v$. Note that if ω_l is a pole on the real axis, $-\omega_l$ is also a pole. Additionally, $\text{Res}(f(\omega))_{\omega=\omega_l} = -\text{Res}(f(\omega))_{\omega=-\omega_l}^*$, so that if $\omega_l = 0$ the residue has vanishing real part.

Building on equation (4.1), the macroscopic susceptibility can then be written as:

$$\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(\omega) = \Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega)\Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega) = \Pi_{ij}^h(\omega)\Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega) + \sum_{l=1}^R \frac{\text{Res}(\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega))_{\omega=\omega_l}}{\omega - \omega_l} \Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega_l), \quad (4.2)$$

where the Kramers–Kronig relations given in equation (3.3) are then modified as follows, taking into account the extra poles along the real ω -axis:

$$\mathbf{P} \int_{-\infty}^{\infty} dv \frac{\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(v)}{v - \omega} = i\pi \tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(\omega) + i\pi \sum_{l=1}^R \frac{\text{Res}(\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega))_{\omega=\omega_l}}{\omega_l - \omega} \Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega_l). \quad (4.3)$$

By taking the limit $\omega \rightarrow \infty$ we can generalize the sum rule given in equation (3.7):

$$\int_0^{\infty} dv \mathbf{Re}\{\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(v)\} = \frac{\pi}{2} \langle X_i(\mathbf{x}) \rangle_0 - \frac{\pi}{2} \mathbf{Im} \left\{ \sum_{l=1}^R \text{Res}(\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega))_{\omega=\omega_l} \Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega_l) \right\}. \quad (4.4)$$

Instead, by taking the limit $\omega \rightarrow 0$ we can generalize the sum rule given in equation (3.8) as follows:

$$\int_0^{\infty} dv \frac{\mathbf{Im}\{\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(v)\}}{v} = \lim_{\omega \rightarrow 0} \left(\frac{\pi}{2} \mathbf{Re}\{\tilde{\Gamma}_{i,\tilde{\alpha},\tilde{\beta}}(\omega)\} \right) + \frac{\pi}{2} \mathbf{Re} \left\{ \sum_{\omega_l \neq 0} \frac{\text{Res}(\Pi_{ij,\tilde{\alpha},\tilde{\beta}}(\omega))_{\omega=\omega_l}}{\omega_l} \Gamma_{i,\tilde{\alpha},\tilde{\beta}}(\omega_l) \right\}, \quad (4.5)$$

where we note that the zero-frequency poles do not contribute to the second term on the right-hand side.

(b) Two scenarios of phase transition

In the discussion above, we are assuming that for $(\alpha, \theta) = (\tilde{\alpha}, \tilde{\theta})$ we have that $\det(P_{ij,\alpha,\beta}(\omega))$ vanishes for R real values of ω . Since $P_{ij,\alpha,\beta}(\omega) = (P_{ij,\alpha,\beta}(-\omega^*))^*$, we have that $\det(P_{ij,\alpha,\beta}(\omega)) = (\det(P_{ij,\alpha,\beta}(-\omega^*)))^*$. Therefore, the solutions to the equation $\det(P_{ij,\alpha,\beta}(\omega)) = 0$ come in conjugate

pairs if they are complex. Generically, we can assume that as we tune the parameter s to the critical value \tilde{s} such that $(\alpha_{\tilde{s}}, \theta_{\tilde{s}}) = (\tilde{\alpha}, \tilde{\theta})$ either one real solution or the real part of one pair of solutions crosses to positive values. We then consider the following two scenarios for the poles $\omega_l, l = 1, \dots, R$:

- $\omega_1 = 0, R = 1$; or
- $\omega_1 = -\omega_2 > 0, R = 2$.

Indeed, we wish to consider the two qualitatively different cases of either (i) a single pole with zero frequency; or (ii) a pair of poles with non-vanishing and opposite frequencies emerging at $(\alpha, \theta) = (\tilde{\alpha}, \tilde{\theta})$. Of course, more than two poles could simultaneously emerge $(\alpha, \theta) = (\tilde{\alpha}, \tilde{\theta})$, but we consider this as a non-generic case.

- If $\omega_l = 0$ is a pole, then we have a static phase transition, associated with a breakdown in the linear response describing the parametric modulation of the measure of the system, see §6a. While such a statement applies for rather general systems and perturbations, this situation can be better understood by considering the specific perturbation $\mathbf{X}(\mathbf{x}) = \langle \mathbf{x} \rangle_0 - \mathbf{x}$ with $T(t) = 1$, which amounts to studying, within linear approximation, how the measure of the system changes as the value of θ is changed to $\theta + \epsilon$. This phase transition corresponds to a insulator-metal phase transition in condensed matter, because the electric susceptibility $\chi_{ij}^{(1)}(\omega)$ of a conductor diverges as $i\sigma_{ij}/\omega$ for small frequencies, where σ is a real tensor and describes the static electric conductivity, which is vanishing for an insulator [30].
- If, instead, we have a pair of poles located at $\pm\omega_l \neq 0$, we have a dynamic phase transition activated by a forcing with non-vanishing spectral power at the frequency $\pm\omega_l$. In this case, a limit cycle emerges corresponding to self-sustained oscillation, which is made possible by the feedback encoded in the nonlinearity of the McKean–Vlasov equation, see e.g. [89] and §6b.

In §6, we will present examples of phase transitions occurring according to the two scenarios above.

5. Equilibrium phase transitions: gradient systems

When the local force can be written as a gradient of a potential $\mathbf{F}_\alpha(\mathbf{y}) = -\nabla V_\alpha(\mathbf{y})$ and the diffusion matrix is the identity matrix $s_{ij} = \delta_{ij}$, equations (2.1) describe an equilibrium system. In particular, the N particles system has a unique ergodic invariant measure when the potential satisfies suitable confining properties [16,63] (see later discussion). Equivalently, the generator of the finite particle stochastic process has purely discrete spectrum, a non-zero spectral gap and the system converges exponentially fast to the unique equilibrium state, both in the L^2 space weighted by the invariant measure and in relative entropy.

In the limit $N \rightarrow \infty$, the system is described by the McKean–Vlasov equation (2.2) whose stationary measures are solutions of the Kirkwood–Monroe equation [105]:

$$\rho_{\alpha,\theta}^{(0)} = \frac{1}{Z} e^{-\langle 2/\sigma^2 \rangle (V(\mathbf{x}) + U * \rho_{\alpha,\theta}^{(0)}(\mathbf{x}))}, \quad Z = \int e^{-\langle 2/\sigma^2 \rangle (V(\mathbf{x}) + U * \rho_{\alpha,\theta}^{(0)}(\mathbf{x}))} \mathrm{d}\mathbf{x}. \quad (5.1)$$

When the confining and interaction potentials are strongly convex and convex, respectively, then it is well known that equation (5.1) has only one solution, corresponding to the unique steady state of the McKean–Vlasov dynamics [106]. In addition, the dynamics converges exponentially fast, in relative entropy, to the stationary state and the rate of convergence to equilibrium can be quantified [106]. However, when the confining potential is not convex, e.g. is bistable, then more than one stationary states can exist, at sufficiently low noise strength (equivalently, for sufficiently strong interactions). A well-known example where the non-uniqueness of the invariant measure

is that of the Desai–Zwanzig model [1,86,103], where the interaction potential is quadratic (see §6a for more details). In this framework, the loss of uniqueness of the invariant measure can be interpreted as a continuous-phase transition, similar to some extent to the phase transition for the Ising model. For a quadratic interaction potential, the equilibrium stationary measure (5.1) can be written as

$$\rho_{\alpha,\theta}^{(0)} = \frac{1}{Z} e^{-(2/\sigma^2)\hat{V}}, \quad Z = \int e^{-(2/\sigma^2)\hat{V}} d\mathbf{x}, \quad (5.2)$$

where we have introduced the modified potential $\hat{V}(\mathbf{x}) = V(\mathbf{x}) - \theta(|\mathbf{x}|^2/2) - \langle \mathbf{x} \rangle_0 \cdot \mathbf{x}$, with the term proportional to θ arising from the interactions between the subsystems. The linear Fokker–Planck operator associated to the stationary Mc-Kean Vlasov equation (2.2) describing the equilibrium dynamics relative to (5.2) reads

$$M_{\alpha,\theta,(\mathbf{x})_0}^0(\cdot) = \nabla \cdot (\nabla \hat{V}(\mathbf{x}) \cdot) + \frac{\sigma^2}{2} \Delta. \quad (5.3)$$

It is well known [63, Sect. 4.5] that, if the modified potential \hat{V} satisfies the property

$$\lim_{|\mathbf{x}| \rightarrow +\infty} \left(\frac{|\nabla \hat{V}|^2}{2} - \Delta \hat{V} \right) = +\infty \quad (5.4)$$

then the operator $M_{\alpha,\theta,(\mathbf{x})_0}^0$ in (5.3) has a spectral gap in $L^2(\rho_{\alpha,\theta}^{(0)})$, the space of square integrable functions weighted with by the invariant density $\rho_{\alpha,\theta}^{(0)}$. In particular, condition (5.4) prevents the system from undergoing a phase transition via scenario (i). When detailed balance holds, the mean field susceptibility $G_{i,\alpha,\theta}(\tau)$ relative to a uniform spatial forcing $\mathbf{X} = \text{const.}$ can be written as the time derivative of suitable correlation functions. In fact, from equation (2.11), the mean field susceptibility can be written as

$$G_{i,\alpha,\theta}(\tau) = -\Theta(\tau) \int d^M \mathbf{y} y_i \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^0 \tau \right) \nabla \cdot \left(\rho_{\alpha,\theta}^{(0)}(\mathbf{y}) \mathbf{X}(\mathbf{y}) \right). \quad (5.5)$$

Without loss of generality, let us consider a uniform forcing $\mathbf{X} = \hat{\mathbf{v}}_k$, with $\hat{\mathbf{v}}_k$ being the unit vector in the k th direction. The mean field susceptibility thus becomes

$$G_{i,\alpha,\theta}(\tau) = -\Theta(\tau) \int d^M \mathbf{y} y_i \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^0 \tau \right) \partial_{y_k} \rho_{\alpha,\theta}^{(0)} = \frac{Y_{\{i,k\},\alpha,\theta}(\tau)}{\theta}. \quad (5.6)$$

Since the system is at equilibrium and the stationary probability density can be written as in (5.1), $\partial_{y_k} \rho_{\alpha,\theta}^{(0)} = -(2/\sigma^2) \rho_{\alpha,\theta}^{(0)} \partial_{y_k} \hat{V}$, physically representing the fact that the probability current associated with the invariant measure vanishes at equilibrium. Furthermore, using (5.3) it is easy to verify the following identity $M_{\alpha,\theta,(\mathbf{x})_0}^0(y_k \rho_{\alpha,\theta}^{(0)}) = -\rho_{\alpha,\theta}^{(0)} \partial_{y_k} \hat{V}$. The mean field susceptibility can then be written as

$$G_{i,\alpha,\theta}(\tau) = -\frac{2}{\sigma^2} \Theta(\tau) \int d^M \mathbf{y} y_i \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^0 \tau \right) M_{\alpha,\theta,(\mathbf{x})_0}^0 y_k \rho_{\alpha,\theta}^{(0)} \quad (5.7)$$

$$= -\frac{2}{\sigma^2} \Theta(\tau) \frac{d}{d\tau} \int d^M \mathbf{y} y_i \exp \left(M_{\alpha,\theta,(\mathbf{x})_0}^0 \tau \right) y_k \rho_{\alpha,\theta}^{(0)} \quad (5.8)$$

$$= -\frac{2}{\sigma^2} \Theta(\tau) \frac{d}{d\tau} \langle x_i(\tau) x_k(0) \rangle_0 \quad (5.9)$$

$$= -\frac{2}{\sigma^2} \Theta(\tau) \frac{d}{d\tau} \langle z_i(\tau) z_k(0) \rangle_0, \quad (5.10)$$

where in the last equation we have introduced the fluctuation variables $z_i = x_i - \langle x_i \rangle_0$. Equation (5.10) shows that the mean field susceptibility is closely related to equilibrium

correlation functions. It is then possible to associate to each correlation function the correlation time

$$\tau_{ij} = \frac{\int_0^{+\infty} dt \langle z_i(\tau) z_j(0) \rangle_0}{\langle z_i(0) z_j(0) \rangle_0}. \quad (5.11)$$

Note that this time scale differs from the one introduced in equation (3.8), which in this case can be written as

$$\tau_{G_i} = \frac{\int_0^{\infty} dt G_{i,\alpha,\beta}(t)}{G_{i,\alpha,\beta}(0^+)} = - \frac{\langle z_i(0) z_j(0) \rangle_0}{\lim_{t \rightarrow 0^+} d/dt \langle z_i(t) z_j(0) \rangle_0}.$$

By comparing the expressions of τ_{G_i} and τ_{ij} and by considering equation (2.15), one understands that τ_{G_i} and τ_{ij} correspond to two differently weighted averages of the timescales associated with each subdominant mode of the operator $M_{\alpha,\beta}^0$.

Usually, the singular behaviour of correlation properties has been used as an indicator of critical transitions [74]. However, let us remark again that, being related to the spectrum of the operator $M_{\alpha,\beta}^0$, in our case neither τ_{G_i} nor τ_{ij} show any critical behaviour at transitions occurring according to the scenario (ii), while they both diverge in the case of critical transitions corresponding to the scenario (i).

6. Examples

In what follows we re-examine the linear response of two relevant models that have been extensively investigated in the literature. Using the framework developed above, we investigate the phase transitions occurring in the Desai–Zwanzig model [86] and the Bonilla–Casado–Morilla model [89], which are taken as paradigmatic examples of equilibrium and non-equilibrium systems, respectively. We also provide the result of numerical simulations for both models.

(a) Equilibrium phase transition: the Desai–Zwanzig model

The Desai–Zwanzig model [86] has a paradigmatic value as it features an equilibrium thermodynamic phase transition (pitchfork bifurcation) arising from the interaction between systems [107] and has been used also as a model for systemic risk [3]. Each of the systems can be interpreted as a particle, moving in one dimension ($M = 1$) in a double well potential $V_\alpha(x) = -(\alpha/2)x^2 + x^4/4$, interacting with the other particles via a quadratic interaction $U(x)$. The N -particle system is described by

$$dx^k = F_\alpha(x^k)dt - \frac{\theta}{N} \sum_{l=1}^N \partial_{x^k} U(x^k - x^l)dt + \sigma dW^k, \quad (6.1)$$

where $k = 1, \dots, N$. The local force is $F_\alpha = -V'_\alpha$, the interaction potential is $U(x) = x^2/2$ and the volatility matrix is the identity matrix $s_{ij} = \delta_{ij}$. Furthermore, V_α is double-well-shaped when $\alpha > 0$, otherwise it has a unique global minimum. In the thermodynamic limit $N \rightarrow \infty$, the one particle density satisfies the McKean–Vlasov equation (2.2) and it has been proven [1,103] that the infinite particle system undergoes a continuous phase transition, with $\langle x \rangle$ being a suitable order parameter. The Desai–Zwanzig model can be seen as a stochastic model of key importance for elucidating order–disorder phase transitions [107].

We have studied the Desai–Zwanzig model also through numerical integration of equation (6.1) by adopting an Euler–Maruyama scheme [108]. We have tested the convergence of our results in the thermodynamic limit $N \rightarrow \infty$ by looking at increasing values of the number N of particles. We present in figures 1a–c the results obtained with $N = 5000$ for $0.2 \leq \theta \leq 1.0$ and $0.4 \leq \sigma \leq 1.0$. The relevant expectation values and correlations have been evaluated considering averages performed over 2.5×10^3 time units. Figure 2a,b portrays two sections performed approximately in the middle of the domain of the heat maps provided in figure 1a–c, with the goal of clarifying the obtained results. The order parameter clearly indicates a continuous-phase transition. The re-scaled variance of the fluctuations, being related to the operator $M_{\alpha,\beta}^0$, is

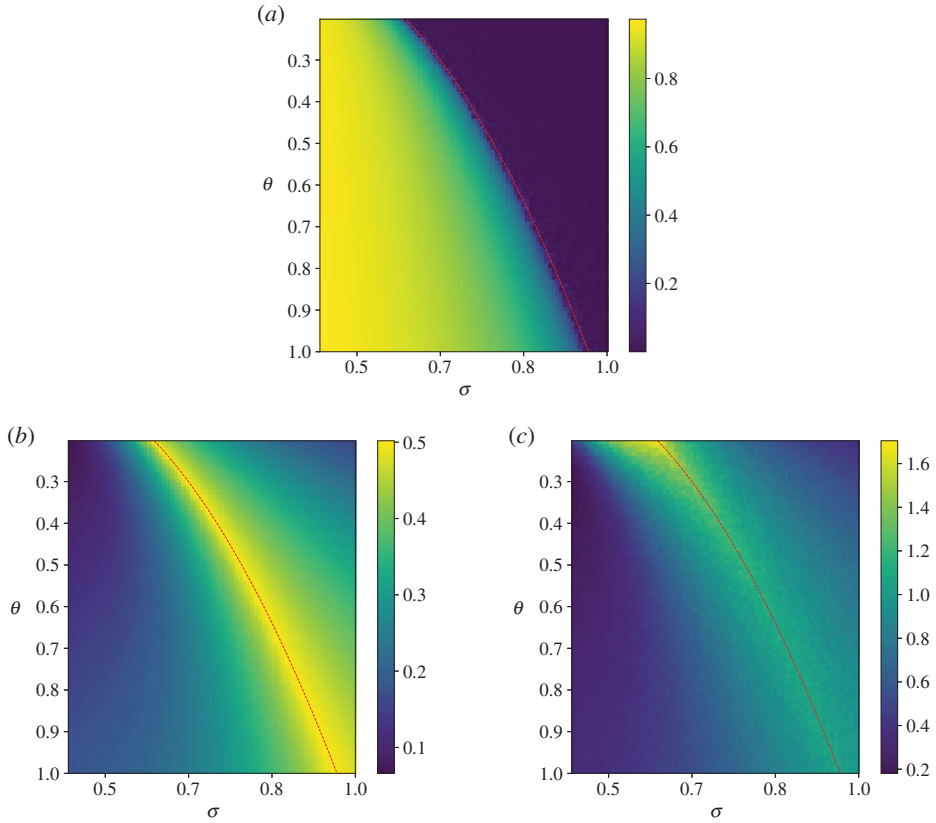


Figure 1. Results of numerical simulations of equation (6.1) with $\alpha = 1$. Heat maps of the order parameter $\langle x \rangle_0$ (panel (a)); of the re-scaled variance $(\theta/\sigma^2)\langle z^2 \rangle_0$ (panel (b)); and of the re-scaled correlation time $\hat{\tau} = \theta \times \tau$ (panel (c)). The dotted red line shows the transition line, see [1]. See text for details. (Online version in colour.)

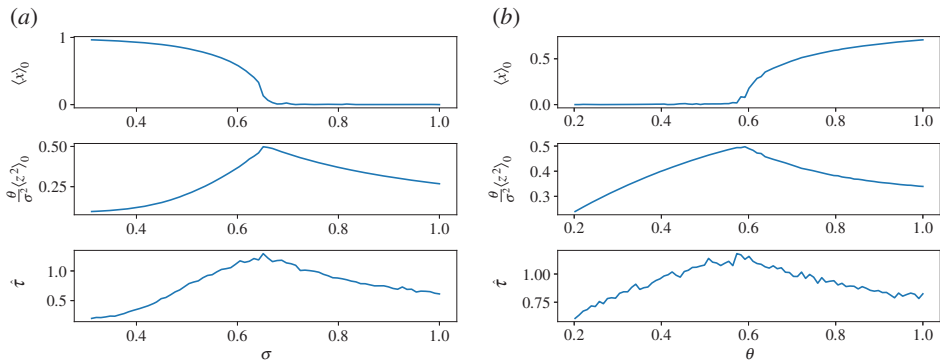


Figure 2. A horizontal (left) and a vertical (right) section of the heat maps shown in figure 1a–c. (a) From top to bottom: order parameter, re-scaled variance and re-scaled integrated autocorrelated time as a function of the strength of the noise. Here $\theta \approx 0.4$. (b) From top to bottom: order parameter, re-scaled variance and re-scaled integrated autocorrelated time as a function of the strength of the coupling. Here $\sigma \approx 0.78$. (Online version in colour.)

finite (and equal to 1/2) at the transition point, in agreement with equation (6.8). The re-scaled correlation time $\hat{\tau} = \theta \times \tau$, where τ is defined in (5.11), is also non-singular, as discussed below. The response of the order parameter to a perturbation $F_\alpha \rightarrow F_\alpha + \varepsilon X(x)T(t)$ is given by

equation (2.8). Given the simplicity of this model, it is possible to explicitly evaluate all the relevant quantities that characterize a phase transition relative to scenario (ii). Indeed, if we consider a perturbation such that $X(x) = 1$, one has $Y_{\alpha,\beta}(\tau) = \theta G_{\alpha,\beta}(\tau)$ and equation (2.16) can be written as

$$P(\omega)\langle x \rangle_1(\omega) = \Gamma_{\alpha,\beta}(\omega)T(\omega) \quad (6.2)$$

where the 1×1 matrix is $P(\omega) = 1 - \theta \Gamma_{\alpha,\beta}(\omega)$. The macroscopic susceptibility is then obtained as

$$\tilde{\Gamma}_{\alpha,\beta}(\omega) = P^{-1}(\omega)\Gamma_{\alpha,\beta}(\omega) = \frac{\Gamma_{\alpha,\beta}(\omega)}{1 - \theta \Gamma_{\alpha,\beta}(\omega)}. \quad (6.3)$$

Furthermore, this is a gradient system satisfying all the assumptions that have been made in §5, so that the mean field susceptibility can be written as (see also [103])

$$G_{\alpha,\beta}(\tau) = -\Theta(\tau) \frac{2}{\sigma^2} \frac{d}{d\tau} \langle z(\tau)z(0) \rangle_0, \quad (6.4)$$

where $z(t) = x - \langle x \rangle_0$. Taking the Fourier transform results in

$$\Gamma_{\alpha,\beta}(\omega) = \frac{2}{\sigma^2} \left[\langle z^2 \rangle_0 - i\omega\gamma(\omega) \right]. \quad (6.5)$$

where $\gamma(\omega) = \int_0^\infty dt e^{-i\omega t} \langle z(t)z(0) \rangle_0$ is the Fourier transform of the correlation function. As previously mentioned, $\Gamma_{\alpha,\beta}(\omega)$ can be written in terms of the spectrum of the operator $M_{\alpha,\beta,\langle x \rangle_0}^{0,+}$ which in this specific example reads (see equation (5.3))

$$M_{\alpha,\beta,\langle x \rangle_0}^{0,+} = -\hat{V}'(x)\partial_x + \frac{\sigma^2}{2} \partial_{xx}, \quad (6.6)$$

where the modified potential is $\hat{V} = V_\alpha - \theta(x^2/2 - \langle x \rangle_0 x)$. It can be proven [1] that the above operator is self-adjoint and has a pure point spectrum $\{\lambda_\mu\}$ with $0 = \lambda_0 > \lambda_1 > \lambda_2 > \dots$, with the vanishing eigenvalue corresponding to the stationary distribution $\rho_{\alpha,\beta}^{(0)}$. In fact, it is easy to show that condition (5.4) holds. The operator $\tilde{M}_{\alpha,\beta,\langle x \rangle_0}^{0,+}$ is instead

$$\tilde{M}_{\alpha,\beta,\langle x \rangle_0}^{0,+}(\rho_{\alpha,\beta}^{(1)}) = M_{\alpha,\beta,\langle x \rangle_0}^{0,+}(\rho_{\alpha,\beta}^{(1)}) - \theta \langle x \rangle_1(t) \partial_x \rho_{\alpha,\beta}^{(0)}. \quad (6.7)$$

Dawson [1] proved that, away from the transition point—in particular, above it, where $\langle x \rangle_0 = 0$ —the nonlinear operator $\tilde{M}_{\alpha,\beta,\langle x \rangle_0}^{0,+}$ has similar spectral properties to $M_{\alpha,\beta,\langle x \rangle_0}^{0,+}$. At the transition, though, $\tilde{M}_{\alpha,\beta,\langle x \rangle_0}^{0,+}$ shows a vanishing spectral gap, with the operator developing a null eigenvalue. This situation corresponds to the breakdown of the aforementioned condition (ii) in which the mean field susceptibility $\Gamma_{\alpha,\beta}(\omega)$ —and thus $\gamma(\omega)$ —is holomorphic in the upper complex ω -plane, while the macroscopic $\tilde{\Gamma}_{\alpha,\beta}$ develops a pole, arising from the non-invertibility of $P(\omega)$. Let us observe again that this implies that at the transition there is no divergence of the integrated autocorrelation time τ , because the spectral gap of the operator $M_{\alpha,\beta,\langle x \rangle_0}^{0,+}$ does not shrink to zero. This is clearly shown in the two-dimensional map shown in figure 1c and in the two sections shown in figure 2a,b. We can fully characterize the singular behaviour of the macroscopic susceptibility $\tilde{\Gamma}_{\alpha,\beta}$ at the transition. As a matter of fact, the transition point is characterized [103] by the condition

$$1 - \frac{2\theta}{\sigma^2} \langle z^2 \rangle_0 = 0 \quad (6.8)$$

so that the macroscopic susceptibility becomes

$$\tilde{\Gamma}_{\alpha,\beta}(\omega) = \frac{(2/\sigma^2) [\langle z^2 \rangle_0 - i\omega\gamma(\omega)]}{i\theta\omega\gamma(\omega)} = -\frac{2}{\theta\sigma^2} + \frac{2\langle z^2 \rangle_0}{i\theta\omega\gamma(\omega)} \quad (6.9)$$

As previously discussed in relation to equation (4.2), the above expression shows that at the transition point $\tilde{\Gamma}_{\alpha,\beta}$ develops a simple pole in $\omega = 0$, with residue

$$\text{Res}(\tilde{\Gamma}_{\alpha,\beta})_{\omega=0} = -i \frac{2\langle z^2 \rangle_0}{\theta\gamma(0)} \quad (6.10)$$

(b) Non-equilibrium phase transition: the Bonilla–Casado–Morilla model

In this section, we will study the Bonilla–Casado–Morrillo model [89] and elucidate the properties of a non-equilibrium self-synchronization phase transition, by looking at the divergence of the macroscopic susceptibility $\tilde{T}_{\alpha,\theta}$. We anticipate that the susceptibility develops a pair of symmetric poles $\omega_1 = -\omega_2 > 0$ at the transition point, thus following the scenario (ii) discussed above. The model consists of N two-dimensional nonlinear oscillators $\mathbf{x}^k = (x_1^k, x_2^k)$, interacting via a quadratic interaction potential $U(\mathbf{x}) = |\mathbf{x}|^2/2$ and subjected to thermal noise

$$d\mathbf{x}_i^k = F_{i,\alpha}(\mathbf{x}^k)dt - \frac{\theta}{N} \sum_{l=1}^N \partial_{\mathbf{x}_i^k} U(\mathbf{x}^k - \mathbf{x}^l)dt + \sigma dW_i^k, \quad k = 1, \dots, N. \quad (6.11)$$

The local force is not conservative, giving rise to a non-equilibrium process, and reads $\mathbf{F}_\alpha(\mathbf{x}) = (\alpha - |\mathbf{x}|^2)\mathbf{x} + \mathbf{x}^+$ where $\mathbf{x}^+ = (-x_2, x_1)$. This term corresponds to a rotation, which is divergence-free with respect to the (Gibbsian) invariant measure and, therefore, does not change the stationary state, but it makes it a non-equilibrium one [109–111]. The systematic study of linear response theory for such non-equilibrium systems is an interesting problem that we leave for future study. In the thermodynamic limit, the system is described by a McKean–Vlasov equation

$$\partial_t \rho(\mathbf{x}, t) = -\nabla \cdot \left[(\hat{\mathbf{F}} + \theta(\mathbf{x})) \rho \right] + \frac{\sigma^2}{2} \Delta \rho, \quad (6.12)$$

where $\hat{\mathbf{F}} = \mathbf{F}_\alpha - \theta\mathbf{x}$, the last term representing the mean field contribution of the coupling to the local force. The authors in [89] prove that the infinite particle system undergoes a phase transition, with a stationary measure $\rho_0(\mathbf{x})$ losing stability to a time-dependent probability measure $\bar{\rho} = \bar{\rho}(\mathbf{x}, t)$. Physically, this phenomenon can be interpreted as a process of synchronization. In fact, $\rho^{(0)}(\mathbf{x})$ represents a disordered state, with the oscillators moving out of phase, while $\bar{\rho}$ describes a state of collective organization with the oscillators moving in an organized rhythmic manner. The transition can be investigated via the order parameter $\langle \mathbf{x} \rangle$ which vanishes in the asynchronous state, $\langle \mathbf{x} \rangle_0 = 0$, and is different from zero and time dependent in the synchronous state. In particular, the stationary measure $\rho_0(\mathbf{x})$ can be written as

$$\rho^{(0)}(\mathbf{x}) = \frac{1}{Z} e^{-\phi(\mathbf{x})}, \quad \phi(\mathbf{x}) = \left(\theta - \alpha + \frac{1}{2} |\mathbf{x}|^2 \right) \frac{|\mathbf{x}|^2}{\sigma^2} \quad (6.13)$$

and satisfies the stationary McKean–Vlasov equation

$$M_{\alpha,\theta,\langle \mathbf{x} \rangle_0}^0 \left(\rho^{(0)} \right) = 0 \quad (6.14)$$

with $M_{\alpha,\theta,\langle \mathbf{x} \rangle_0}^0(g) = -\nabla \cdot [\hat{\mathbf{F}}g] + (\sigma^2/2)\Delta g$, being the Fokker–Planck operator describing the stationary state $\rho^{(0)}(\mathbf{x})$. Note that $\langle \mathbf{x} \rangle_0 = 0$. We can perform a linear response theory around this stationary state ρ_0 by replacing $\mathbf{F}_\alpha \rightarrow \mathbf{F}_\alpha + \varepsilon \mathbf{X}(\mathbf{x})T(t)$ and studying the perturbation $\rho^{(1)}$ of the measure defined via $\rho(\mathbf{x}, t) = \rho^{(0)}(\mathbf{x}) + \varepsilon \rho^{(1)}(\mathbf{x}, t)$. As previously outlined, $\rho^{(1)}(\mathbf{x}, t)$ satisfies equation (2.4) from which the whole linear response theory follows. However, to conform to the notation in [89] we will here define $\rho^{(1)}(\mathbf{x}, t) = (\rho^{(0)})^{1/2} q(\mathbf{x}, t)$ and write the corresponding equation for $q(\mathbf{x}, t)$. After some algebra, it is possible to write that

$$\begin{aligned} \partial_t q(\mathbf{x}, t) &= \mathcal{M}_{\alpha,\theta,0}(q) - T(t)(\rho^{(0)})^{-1/2} \nabla \cdot \left(\mathbf{X}(\mathbf{x})\rho^{(0)} \right) + \theta(\rho^{(0)})^{1/2} \langle (\rho^{(0)})^{1/2} \mathbf{y}, q(\mathbf{y}, t) \rangle \cdot \nabla \phi(\mathbf{x}) \\ &= \tilde{\mathcal{M}}_{\alpha,\theta,0}(q) - T(t)(\rho^{(0)})^{-1/2} \nabla \cdot \left(\mathbf{X}(\mathbf{x})\rho^{(0)} \right) \end{aligned} \quad (6.15)$$

where we have defined

$$\mathcal{M}_{\alpha,\theta,0}(q) = \frac{\sigma^2}{4} \left[\Delta \phi - \frac{1}{2} |\nabla \phi|^2 \right] q + [-\mathbf{x}^+ \cdot \nabla + \Delta] q \quad (6.16)$$

We mention that operator has the structure of a Schrödinger operator in a magnetic field ([63], Sec. 4.9). Furthermore, $\tilde{\mathcal{M}}_{\alpha,\theta,0}(q) = \mathcal{M}_{\alpha,\theta,0}(q) + \theta(\rho^{(0)})^{1/2} \langle (\rho^{(0)})^{1/2} \mathbf{y}, q(\mathbf{y}, t) \rangle \cdot \nabla \phi(\mathbf{x})$ with $\langle f, g \rangle = \int dy f(\mathbf{y})g(\mathbf{y})$ being the usual scalar product. In particular, let us observe that

$\langle (\rho^{(0)})^{1/2} \mathbf{y}, q(\mathbf{y}, t) \rangle = \int (\rho^{(0)})^{1/2} \mathbf{y} q(\mathbf{y}, t) d\mathbf{y} = \int \mathbf{y} \rho^{(1)}(\mathbf{y}, t) d\mathbf{y} = \langle \mathbf{y} \rangle_1$. A formal solution of the above equation is

$$q(\mathbf{x}, t) = \int_{-\infty}^t ds \exp[\mathcal{M}_{\alpha, \beta, 0}(t-s)] \left[-T(s)(\rho^{(0)})^{-1/2} \nabla \cdot (\mathbf{X}(\mathbf{x}) \rho^{(0)}) + \theta (\rho^{(0)})^{1/2} \langle (\rho^{(0)})^{1/2} \mathbf{y}, q(\mathbf{y}, s) \rangle \cdot \nabla \phi(\mathbf{x}) \right] \quad (6.17)$$

which is the analogous of equation (2.5). Using the above expression, we can evaluate the response of the observable x_i as

$$\langle x_i \rangle_1 = \left\langle (\rho^{(0)})^{1/2} x_i, q(\mathbf{x}, t) \right\rangle = \int d\mathbf{x} \int_{-\infty}^t ds \quad x_i \exp[\mathcal{M}_{\alpha, \beta, 0}(t-s)] \times \left[-T(s)(\rho^{(0)})^{-1/2} \nabla \cdot (\mathbf{X}(\mathbf{x}) \rho^{(0)}) + \theta (\rho^{(0)})^{1/2} \langle (\rho^{(0)})^{1/2} \mathbf{y}, q(\mathbf{y}, s) \rangle \cdot \nabla \phi(\mathbf{x}) \right]. \quad (6.18)$$

Comparing equation (6.18) and (2.6), it is clear that the operators $\mathcal{M}_{\alpha, \beta, 0}, \tilde{\mathcal{M}}_{\alpha, \beta, 0}$ are analogous to the operators $M_{\alpha, \beta, \langle \mathbf{x} \rangle_0}^0, \tilde{M}_{\alpha, \beta, \langle \mathbf{x} \rangle_0}^0$ defined in §2. In particular, their spectrum is related to the Fourier transform of the mean field susceptibility $\Gamma_{\alpha, \beta}$ and macroscopic susceptibility $\tilde{\Gamma}_{\alpha, \beta}$ (respectively) through equations similar to (2.17) and (2.27). The authors in [89] study the spectrum of both these operators in order to perform a stability analysis of the stationary distribution $\rho_0(\mathbf{x})$. In particular, they observe that the operator $\mathcal{M}_{\alpha, \beta, 0}$ can be written as $\mathcal{M}_{\alpha, \beta, 0} = \mathcal{M}_H + \mathcal{M}_A$ where

$$\mathcal{M}_H(q) = \frac{\sigma^2}{4} \left[\Delta \phi - \frac{1}{2} |\nabla \phi|^2 \right] q + \frac{\sigma^2}{2} \Delta q \quad (6.19)$$

and

$$\mathcal{M}_A(q) = -\mathbf{x}^+ \cdot \nabla q \quad (6.20)$$

with vanishing commutator $[\mathcal{M}_H, \mathcal{M}_A] = 0$. The operator \mathcal{M}_H is related to the conservative part of the local force. As a matter of fact, it is a self-adjoint (Hermitian) operator with real eigenvalues. \mathcal{M}_A is instead anti-Hermitian, with purely imaginary eigenvalues (describing oscillations) given by the non-conservative part of \mathbf{F} . Furthermore, \mathcal{M}_H has only one zero eigenvalue corresponding to the ground state $(\rho^{(0)})^{1/2}$ while all the remaining eigenvalues are negative, meaning that scenario (i) in §4b, according to which the spectral gap of the mean field operator $M_{\alpha, \beta, \langle \mathbf{x} \rangle_0}^0$ vanishes, cannot happen in this setting. In particular, correlation properties will never diverge. Phase transition can, instead, take place according to the scenario (ii) above. Indeed, the authors in [89] show that the spectral gap of the operator $\tilde{\mathcal{M}}_{\alpha, \beta, 0}$ vanishes at surface in the (α, σ, θ) parametric space defined by the following equation:

$$A = \frac{\delta^2}{2} \left[1 - \frac{1}{\delta} \exp\left(-\frac{A^2}{\delta^2}\right) \left[\int_{-(A/\delta)}^{\infty} e^{-r^2} dr \right]^{-1} \right], \quad (6.21)$$

where $A = \alpha/\theta - 1$ and $\delta = \sqrt{2\sigma^2}/\theta$. In particular, they are able to prove that the eigenvalues associated with eigenfunctions of $\tilde{\mathcal{M}}_{\alpha, \beta, 0}$ which are orthogonal to the subspace of $L^2(\mathbb{R}^2)$ spanned by $(\rho^{(0)})^{1/2}$ and $\mathbf{n} \cdot \mathbf{x}(\rho^{(0)})^{1/2}$, $\mathbf{n} \in \mathbb{R}^2$ being any unit vector, are always negative. Nevertheless, $\tilde{\mathcal{M}}_{\alpha, \beta, 0}$ can become unstable from eigenfunctions which are not orthogonal to $\mathbf{n} \cdot \mathbf{x}(\rho^{(0)})^{1/2}$. It is possible to identify the eigenfunctions that at the transition yield eigenvalues with vanishing real part. In particular, at the transition line (6.21), the eigenfunction $\Omega(\mathbf{x}) = (0, 1) \cdot \mathbf{x}(\rho^{(0)})^{1/2} + i(1, 0) \cdot \mathbf{x}(\rho^{(0)})^{1/2}$ gives an eigenvalue $\tilde{\lambda}_j = i$, with $\Omega(\mathbf{x})^*$ corresponding to the complex conjugate eigenvalue $\tilde{\lambda}_j^* = -i$. The macroscopic susceptibility (2.27) consequently develops a pair of symmetric poles in $\omega = \pm 1$, corresponding to a dynamic phase transition, giving rise to a Hopf-like bifurcation yielding the time-dependent state $\tilde{\rho}(\mathbf{x}, t)$ that defines the synchronized state. As a result, near the transition, the order parameter $\langle x \rangle$, where the expectation value is computed using the measure $\tilde{\rho}(\mathbf{x}, t)$, oscillates at frequency $\omega = 1$ with amplitude $A_1(\alpha, \sigma, \theta)$. Instead, since it is a quadratic quantity, the rescaled variance $\theta/\sigma^2 \langle z^2 \rangle$, where $z = x - \langle x \rangle$, oscillates at frequency $\omega = 2$ with amplitude $A_2(\alpha, \sigma, \theta)$ around the value $B_2(\alpha, \sigma, \theta)$.

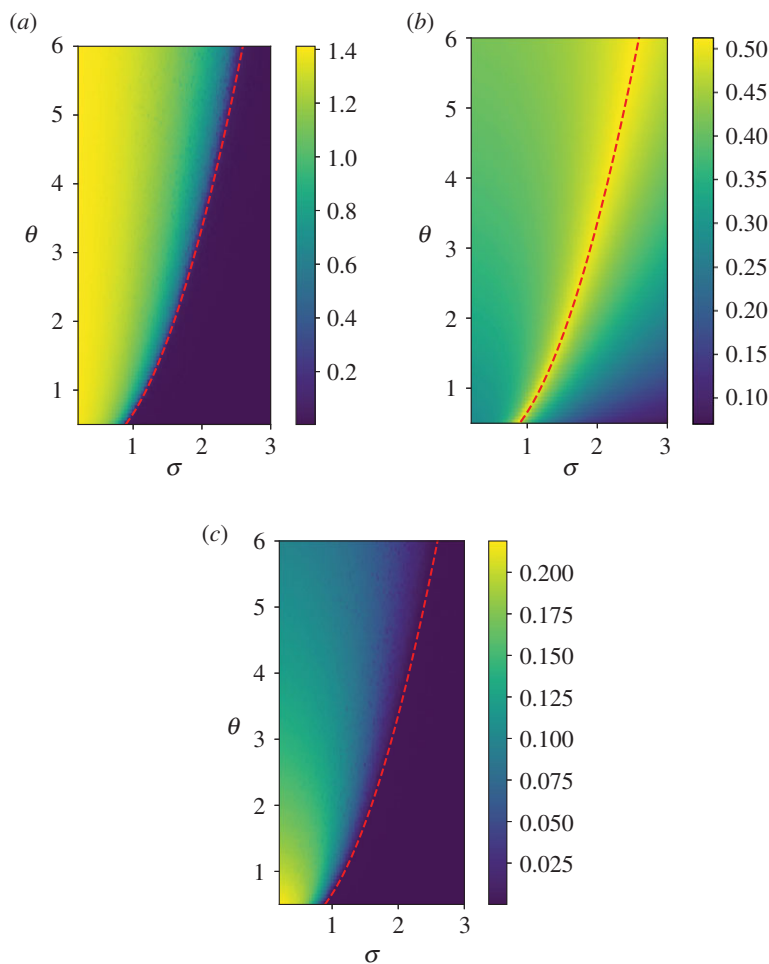


Figure 3. Results of numerical simulations of equation (6.11) with $\alpha = 2$. Heat maps of the amplitude A_1 of the oscillations of the variable x (panel (a)), of the amplitude A_2 of the oscillations of the re-scaled variance $\theta/\sigma^2\langle z^2 \rangle$ (panel (b)), and of the time mean value of $\theta/\sigma^2\langle z^2 \rangle$ (panel (c)). The red dotted line represents the transition line given by equation (6.21); see [89]. See text for details. (Online version in colour.)

We have investigated this non-equilibrium transition through numerical integration of equation (6.11) via an Euler–Maruyama scheme [108]. The convergence of our results to the thermodynamic limit has been tested by looking at increasing values of the number of agents. We display here the results by taking $N = 5000$ and choosing $\alpha = 2$. Figure 3 shows the value of $A_1(\alpha = 2, \sigma, \theta)$ (panel (a)), $A_2(\alpha = 2, \sigma, \theta)$ (panel (b)) and $B_2(\alpha = 2, \sigma, \theta)$ (panel (c)) in the parametric region $0.2 \leq \sigma \leq 3$, $0.5 \leq \theta \leq 6$ of the two dimensional parameter space (σ, θ) . For the sake of clarity, we also provide in figure 4 a snapshot of a horizontal and vertical section of the heat plots.

These numerical experiments confirm that the system indeed undergoes a continuous phase transition, with a collective synchronization stemming from a disordered state as the system passes through the transition line given by equation (6.21) for $\alpha = 2$. Let us remark again that the fluctuations, being related to the spectrum of $M_{\alpha, \theta, (x)_0}^{0,+}$, are always finite, see figure 4.

7. Conclusion

The understanding of how a network of exchangeable interacting systems responds to perturbations is a problem of great relevance in mathematics, natural and social sciences, and

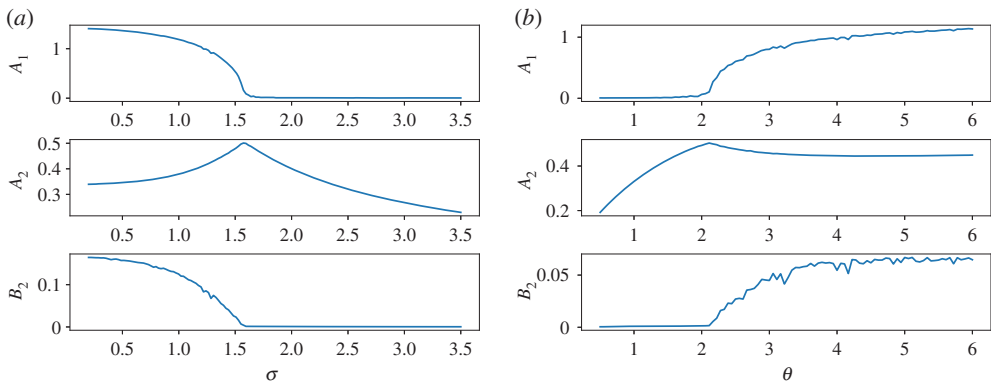


Figure 4. Horizontal (left) and vertical (right) sections of the heat plots $3a$ – c . (a) From top to bottom: $A_1(\alpha = 2, \sigma, \theta = 2)$, $A_2(\alpha = 2, \sigma, \theta = 2)$, and $B_2(\alpha = 2, \sigma, \theta = 2)$. (b) From top to bottom: $A_1(\alpha = 2, \sigma = 1.6, \theta)$, $A_2(\alpha = 2, \sigma = 1.6, \theta)$, and $B_2(\alpha = 2, \sigma = 1.6, \theta)$. (Online version in colour.)

technology. One is in general interested in both the smooth regime of response, where small perturbations result into small changes in the properties of the system, and in the non-smooth regime, which anticipates the occurrence of critical, possibly undesired, changes. Often, critical phenomena, which can be triggered by exogenous or endogenous processes, are accompanied by the existence of a large-scale restructuring of the system, whereby spatial (i.e. across systems) and temporal correlations are greatly enhanced. The emergence of a specific spatial structure is especially clear when considering order–disorder transitions. Spatial–temporal coordination becomes evident when studying the multi-faceted phenomenon of synchronization. Finally, slow decay of temporal correlations—the so-called slowing down—indicates that nearby critical transitions the negative feedback of the system become ineffective.

This paper is the first step in a research programme that aims at developing practical tools for better understanding and predicting—in a data-driven framework—critical transitions in complex systems. We have here developed a fairly general theory of linear response for such a network in the thermodynamic limit of an infinite number of identical interacting systems undergoing deterministic and stochastic forcing. Our approach is able to accommodate both equilibrium and non-equilibrium stationary states, thus going beyond the classical approximation of gradient flows. We remark that the existence of equilibrium stationary (Gibbs) states, the gradient structure (in a suitable metric) and the self-adjointness of the Fokker–Planck operator are equivalent. The presence of interaction between the systems leads to McKean–Vlasov evolution equation for the one-particle density, which reduces to the classical Fokker–Planck equation if the coupling is switched off.

We find explicit expressions for the linear susceptibility and are able to evaluate its asymptotic behaviour, thus allowing for the derivation of a general set of Kramers–Kronig relations and related sum rules. The susceptibility, in close parallel to the classic Clausius–Mossotti expression of macroscopic electric susceptibility for condensed matter, is written in a renormalized form as the product of a matrix describing the self-action of the system times the mean field susceptibility. This allows for further clarifying the relationship between endogenous and exogenous processes, which generalized the fluctuation–dissipation theorem for this class of systems.

Linear response breaks down when the susceptibility diverges, i.e. it develops poles in the real axis. We separate two scenarios of criticality—one associated with the divergence of the mean field susceptibility, and another one associated with singularities of the matrix describing the self-action of the system. The first case pertains to the classical theory of critical transitions.

The second case is here for us of greater interest and can be realized *only* in the thermodynamic limit. We interpret such a second scenario as describing phase transitions for the system. We

define two scenarios of phase transition—a static one, and a dynamic one, where a pole at vanishing frequency and two poles at opposite frequency appear in the linear susceptibility, respectively. At the phase transition, the Kramers–Kronig relations and sum rules valid in the smooth regime of response break down and a detailed study of the poles allows one to find the correction terms. Again, one can establish a link with results from condensed matter physics, as the correction terms resemble those appearing when studying frequency-dependent optical properties of a material at the insulator–metal phase transition, where the static conductivity becomes non-vanishing. We prove that, against intuition, a phase transition is—as opposed to the case of critical transitions—*not* accompanied by a divergence in the autocorrelation properties of the system, i.e. no critical slowing down is observed. Our interpretation is supported by the use of the formalism developed in this paper to revisit through analytical and numerical tools the classical results for phase transitions occurring in the Desai–Zwanzig model on the Bonilla–Casado–Morillo model, for which it is easy to define appropriate order parameters. The criticalities in these two models conform to the scenario of static and dynamic phase transition, respectively.

We remark that studying the linear response of the order parameter is the optimal choice for detecting the phase transition but not the only one. In fact, we expect that a broader class of observables can be used in order to identify the critical behaviour. This is especially important in non-equilibrium cases, where the identification of such order parameter can be extremely non-trivial.

The work reported in this paper opens up several avenues for future research. Four natural next steps are: (a) to investigate in greater detail multi-dimensional reversible (equilibrium) McKean–Vlasov dynamics exhibiting phase transitions; for such systems the self-adjointness of the linearized McKean–Vlasov operator enables the systematic use of tools from spectral theory for selfadjoint operators in appropriate Hilbert spaces. (b) To use the analytical tools developed in this paper to design early warning signals for phase transitions, as opposed to critical transitions for which there exists an extensive literature. (c) To better define the class of observables for which the divergence of the linear response can be used to define and detect phase transitions. In particular, we aim at developing systematic analytical and data-driven methodologies for identifying order parameters in agent based models. These tools will enable us to move beyond the quadratic interaction between subsystems that was considered in this paper. (d) To use the framework developed in this paper in order to revisit phenomena such as synchronization, cooperation and consensus in multi-agent systems, and more generally the emergence of coherent structures in complex systems, both in natural and social sciences as well as technology.

Data accessibility. The codes used to run the simulations and the data used for producing the figures are available on figshare.com at https://figshare.com/projects/Response_theory_and_phase_transition_for_thermodynamic_limit_of_interacting_identical_systems/89516.

Authors' contributions. G.P. initiated the study by formulating the general problem and highlighting the relevance of the Desai–Zwanzig and Bonilla–Casado–Morillo models; V.L. developed most of the theoretical framework; N.Z. investigated the Desai–Zwanzig and Bonilla–Casado–Morillo models, performed the numerical simulations and the related data analysis. All authors contributed to the writing of the paper. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. V.L. 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 G.P. was partially funded by the EPSRC, grant no. EP/P031587/1, and by J.P. Morgan Chase & Co. Any views or opinions expressed herein are solely those of the authors listed, and may differ from the views and opinions expressed by J.P. Morgan Chase & Co. or its affiliates. This material is not a product of the Research Department of J.P. Morgan Securities LLC. This material does not constitute a solicitation or offer in any jurisdiction. N.Z. 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).

Acknowledgements. V.L. wishes to thank A. Pikovsky for his very insightful criticism on the applicability of response theory near the regime of synchronization; and D. Sornette for some useful exchanges on exogenous versus endogenous dynamics in the course of a virtual conference.

References

1. Dawson DA. 1983 Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Stat. Phys.* **31**, 29–85. (doi:10.1007/BF01010922)
2. Acebrón JA, Bonilla LL, Pérez Vicente CJ, Ritort F, Spigler R. 2005 The kuramoto model: a simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* **77**, 137–185. (doi:10.1103/RevModPhys.77.137)
3. Garnier J, Papanicolaou G, Yang T. 2013 Large deviations for a mean field model of systemic risk. *SIAM J. Financ. Math.* **4**, 151–184. (doi:10.1137/12087387X)
4. Wang C, Li Q, B. Chazelle WE. 2017 Noisy Hegselmann-Krause systems: phase transition and the 2r-conjecture. *J. Stat. Phys.* **166**, 1209–1225. (doi:10.1007/s10955-017-1718-x)
5. Garnier J, Papanicolaou G, Yang T. 2017 Consensus convergence with stochastic effects. *Vietnam J. Math.* **45**, 51–75. (doi:10.1007/s10013-016-0190-2)
6. Simmonds J, Gómez JA, Ledezma A. 2019 The role of agent-based modeling and multi-agent systems in flood-based hydrological problems: a brief review. *J. Water and Climate Change* jwc2019108. (doi:10.2166/wcc.2019.108)
7. Geisendorf S. 2018 Evolutionary climate-change modelling: a multi-agent climate-economic model. *Comput. Econ.* **52**, 921–951. (doi:10.1007/s10614-017-9740-2)
8. Pareschi L, Toscani G. 2013 Interacting multiagent systems: kinetic equations and Monte Carlo methods.
9. Kinny D, jen Hsu JY, Governatori G, Ghose AK (eds). 2011 *Agents in Principle, Agents in Practice 14th Int. Conf., PRIMA 2011, Wollongong, Australia, 16–18 November 2011, Proceedings*. Lecture Notes in Artificial Intelligence; 7047, Berlin, Germany: Springer, 1st ed. 2011.
10. Garbuno-Inigo A, Nüsken N, Reich S. 2020 Affine invariant interacting Langevin dynamics for Bayesian inference. *SIAM J. Appl. Dyn. Syst.* **19**, 1633–1658. (doi:10.1137/19M1304891)
11. Oelschläger K. 1984 A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.* **12**, 458–479. (doi:10.1214/aop/1176993301)
12. Gärtner J. 1988 On the McKean–Vlasov limit for interacting diffusions. *Math. Nachr.* **137**, 197–248. (doi:10.1002/mana.19881370116)
13. Frank TD. 2005 *Nonlinear Fokker-Planck equations fundamentals and applications*. Berlin, Germany: Springer.
14. Shiino M. 1987 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**, 2393–2412. (doi:10.1103/PhysRevA.36.2393)
15. Chayes L, Panferov V. 2010 The McKean–Vlasov equation in finite volume. *J. Stat. Phys.* **138**, 351–380. (doi:10.1007/s10955-009-9913-z)
16. Tamura Y. 1984 On asymptotic behaviors of the solution of a nonlinear diffusion equation. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **31**, 195–221.
17. Pikovsky A, Kurths J, Rosenblum M, Kurths J. 2003 *Synchronization: a universal concept in nonlinear sciences*. Cambridge Nonlinear Science Series. Cambridge, UK: Cambridge University Press.
18. Carrillo JA, Gvalani RS, Pavliotis GA, Schlichting A. 2020 Long-time behaviour and phase transitions for the McKean–Vlasov equation on the torus. *Arch. Ration. Mech. Anal.* **235**, 635–690. (doi:10.1007/s00205-019-01430-4)
19. Delgadino MG, Gvalani RS, Pavliotis GA. 2020 On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions. (<https://arxiv.org/abs/2001.03920> [math.AP])
20. Gomes S, Pavliotis G. 2018 Mean field limits for interacting diffusions in a two-scale potential. *J. Nonlin. Sci.* **28**, 905–941. (doi:10.1007/s00332-017-9433-y)
21. Berglund N, Fernandez B, Gentz B. 2007 Metastability in interacting nonlinear stochastic differential equations: I. From weak coupling to synchronization. *Nonlinearity* **20**, 2551–2581. (doi:10.1088/0951-7715/20/11/006)
22. Berglund N, Fernandez B, Gentz B. 2007 Metastability in interacting nonlinear stochastic differential equations: II. Large-behaviour. *Nonlinearity* **20**, 2583–2614. (doi:10.1088/0951-7715/20/11/007)
23. Stefanski A. 2009 *Determining thresholds of complete synchronization, and application*. World Scientific series on nonlinear science: monographs and treatises. Singapore: World Scientific Publishing Company.

24. Balanov A, Janson N, Postnov D, Sosnovtseva O. 2008 *Synchronization: from simple to complex*. Springer Series in Synergetics. Berlin, Germany: Springer.
25. Femat R, Solis-Perales G. 2008 *Robust synchronization of chaotic systems via feedback*. Lecture Notes in Control and Information Sciences. Berlin, Germany: Springer.
26. Boccaletti S, Pisarchik A, Genio C, Amann A. 2018 *Synchronization: from coupled systems to complex networks*. Cambridge, UK: Cambridge University Press.
27. Boccaletti S, Kurths J, Osipov G, Valladares D, Zhou C. 2002 The synchronization of chaotic systems. *Phys. Rep.* **366**, 1–101. (doi:10.1016/S0370-1573(02)00137-0)
28. Eroglu D, Lamb JSW, Pereira T. 2017 Synchronisation of chaos and its applications. *Contemp. Phys.* **58**, 207–243. (doi:10.1080/00107514.2017.1345844)
29. Kubo R. 1966 The fluctuation–dissipation theorem. *Rep. Prog. Phys.* **29**, 255–284. (doi:10.1088/0034-4885/29/1/306)
30. Lucarini V, Saarinen JJ, Peiponen K-E, Vartiainen EM. 2005 *Kramers-Kronig relations in optical materials research*. New York, NY: Springer.
31. Binney J, Tremaine S. 2008 *Galactic dynamics*, 2nd edn. Princeton, NJ: Princeton University Press.
32. Öttinger H. 2005 *Beyond equilibrium thermodynamics*, Hoboken, NJ: Wiley.
33. Marconi UMB, Puglisi A, Rondoni L, Vulpiani A. 2008 Fluctuation-dissipation: response theory in statistical physics. *Phys. Rep.* **461**, 111–195. (doi:10.1016/j.physrep.2008.02.002)
34. Baiesi M, Maes C. 2013 An update on the nonequilibrium linear response. *New J. Phys.* **15**, 013004. (doi:10.1088/1367-2630/15/1/013004)
35. Cessac B. 2019 Linear response in neuronal networks: from neurons dynamics to collective response. *Chaos* **29**, 103105. (doi:10.1063/1.5111803)
36. Sarracino A, Vulpiani A. 2019 On the fluctuation–dissipation relation in non-equilibrium and non-hamiltonian systems. *Chaos* **29**, 083132. (doi:10.1063/1.5110262)
37. Gottwald GA. 2020 Introduction to focus issue: linear response theory: potentials and limits. *Chaos* **30**, 020401. (doi:10.1063/5.0003135)
38. Lucarini V. 2018 Revising and extending the linear response theory for statistical mechanical systems: evaluating observables as predictors and predictands. *J. Stat. Phys.* **173**, 1698–1721. (doi:10.1007/s10955-018-2151-5)
39. Leith CE. 1975 Climate response and fluctuation dissipation. *J. Atmos. Sci.* **32**, 2022. (doi:10.1175/1520-0469(1975)032<2022:CRAFD>2.0.CO;2)
40. North G, Bell R, Hardin J. 1993 Fluctuation dissipation in a general circulation model. *Clim. Dyn.* **8**, 259–264. (doi:10.1007/BF00209665)
41. Gritsun A, Branstator G, Majda AJ. 2008 Climate response of linear and quadratic functionals using the fluctuation–dissipation theorem. *J. Atmos. Sci.* **65**, 2824–2841. (doi:10.1175/2007JAS2496.1)
42. Lucarini V, Ragone F, Lunkeit F. 2017 Predicting climate change using response theory: global averages and spatial patterns. *J. Stat. Phys.* **166**, 1036–1064. (doi:10.1007/s10955-016-1506-z)
43. Bódai T, Lucarini V, Lunkeit F. 2020 Can we use linear response theory to assess geoengineering strategies? *Chaos* **30**, 023124. (doi:10.1063/1.5122255)
44. Lembo V, Lucarini V, Ragone F. 2020 Beyond forcing scenarios: predicting climate change through response operators in a coupled general circulation model. *Sci. Rep.* **10**, 8668. (doi:10.1038/s41598-020-65297-2)
45. Chekroun MD, Simonnet E, Ghil M. 2011 Stochastic climate dynamics: random attractors and time-dependent invariant measures. *Physica D* **240**, 1685–1700. (doi:10.1016/j.physd.2011.06.005)
46. Ghil M, Lucarini V. 2020 The physics of climate variability and climate change. *Rev. Mod. Phys.* **92**, 035002. (doi:10.1103/RevModPhys.92.035002)
47. Ruelle D. 1998 Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms. *Nonlinearity* **11**, 5–18. (doi:10.1088/0951-7715/11/1/002)
48. Ruelle D. 2009 A review of linear response theory for general differentiable dynamical systems. *Nonlinearity* **22**, 855–870. (doi:10.1088/0951-7715/22/4/009)
49. Gritsun A, Lucarini V. 2017 Fluctuations, response, and resonances in a simple atmospheric model. *Physica D* **349**, 62–76. (doi:10.1016/j.physd.2017.02.015)
50. Baladi V. 2000 *Positive transfer operators and decay of correlations*. Singapore: World Scientific.
51. Liverani C, Gouëzel S. 2006 Banach spaces adapted to Anosov systems. *Ergodic Theory and Dyn. Syst.* **26**, 189–217. (doi:10.1017/S0143385705000374)

52. Butterley O, Liverani C. 2007 Smooth Anosov flows: correlation spectra and stability. *J. Modern Dyn.* **1**, 301–322. (doi:10.3934/jmd.2007.1.301)
53. Baladi V. 2008 Linear response despite critical points. *Nonlinearity* **21**, T81–T90. (doi:10.1088/0951-7715/21/6/T01)
54. Abramov R, Majda A. 2007 Blended response algorithms for linear fluctuation–dissipation for complex nonlinear dynamical systems. *Nonlinearity* **20**, 2793–2821. (doi:10.1088/0951-7715/20/12/004)
55. Falcioni M, Vulpiani A. 1995 The relevance of chaos for the linear response theory. *Physica A* **215**, 481–494. (doi:10.1016/0378-4371(94)00277-Z)
56. Cessac B, Sepulchre J-A. 2007 Linear response, susceptibility and resonances in chaotic toy models. *Physica D* **225**, 13–28. (doi:10.1016/j.physd.2006.09.034)
57. Lucarini V, Sarno S. 2011 A statistical mechanical approach for the computation of the climatic response to general forcings. *Nonlin. Process. Geophys.* **18**, 7–28. (doi:10.5194/npg-18-7-2011)
58. Wang Q. 2013 Forward and adjoint sensitivity computation of chaotic dynamical systems. *J. Comput. Phys.* **235**, 1–13. (doi:10.1016/j.jcp.2012.09.007)
59. Chandramoorthy N, Wang Q. 2020 A computable realization of Ruelle’s formula for linear response of statistics in chaotic systems. (<http://arxiv.org/abs/2002.04117>).
60. Ni A. 2020 Approximating Ruelle’s linear response formula by shadowing methods. (<http://arxiv.org/abs/2003.09801>).
61. Hänggi P, Thomas H. 1982 Stochastic processes: time evolution, symmetries and linear response. *Phys. Rep.* **88**, 207–319. (doi:10.1016/0370-1573(82)90045-X)
62. Risken H. 1989 *The Fokker–Planck equation*. Springer Series in Synergetics, vol. 18. Berlin, Germany: Springer.
63. Pavliotis GA. 2014 *Stochastic processes and applications*, vol. 60. New York, NY: Springer.
64. Hairer M, Majda AJ. 2010 A simple framework to justify linear response theory. *Nonlinearity* **23**, 909–922. (doi:10.1088/0951-7715/23/4/008)
65. Dembo A, Deuschel J-D. 2010 Markovian perturbation, response and fluctuation dissipation theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **46**, 822–852. (doi:10.1214/10-AIHP370)
66. Wormell CL, Gottwald GA. 2019 Linear response for macroscopic observables in high-dimensional systems. *Chaos* **29**, 113127. (doi:10.1063/1.5122740)
67. Frank T. 2004 Fluctuation–dissipation theorems for nonlinear Fokker–Planck equations of the Desai–Zwanzig type and Vlasov–Fokker–Planck equations. *Phys. Lett. A* **329**, 475–485. (doi:10.1016/j.physleta.2004.07.019)
68. Fernandez B, Méléard S. 1997 A hilbertian approach for fluctuations on the McKean–Vlasov model. *Stoch. Process. Appl.* **71**, 33–53. (doi:10.1016/S0304-4149(97)00067-7)
69. Pollicott M. 1985 On the rate of mixing of Axiom A flows. *Inventiones Math.* **81**, 413–426. (doi:10.1007/BF01388579)
70. Ruelle D. 1986 Resonances of chaotic dynamical systems. *Phys. Rev. Lett.* **56**, 405–407. (doi:10.1103/PhysRevLett.56.405)
71. Lucarini V. 2016 Response operators for Markov processes in a finite state space: radius of convergence and link to the response theory for Axiom A systems. *J. Stat. Phys.* **162**, 312–333. (doi:10.1007/s10955-015-1409-4)
72. Chekroun MD, Neelin JD, Kondrashov D, McWilliams JC, Ghil M. 2014 Rough parameter dependence in climate models and the role of Ruelle–Pollicott resonances. *Proc. Natl Acad. Sci. USA* **111**, 1684–1690. (doi:10.1073/pnas.1321816111)
73. Tantet A, Lucarini V, Dijkstra HA. 2018 Resonances in a chaotic attractor crisis of the Lorenz flow. *J. Stat. Phys.* **170**, 584–616. (doi:10.1007/s10955-017-1938-0)
74. Scheffer M. 2009 *Critical transitions in nature and society*. Princeton Studies in Complexity. Princeton, NJ: Princeton University Press.
75. Dakos V, Scheffer M, van Nes EH, Brovkin V, Petoukhov V, Held H. 2008 Slowing down as an early warning signal for abrupt climate change. *Proc. Natl Acad. Sci. USA* **105**, 14 308–14 312. (doi:10.1073/pnas.0802430105)
76. Kuehn C. 2011 A mathematical framework for critical transitions: bifurcations, fast-slow systems and stochastic dynamics. *Physica D* **240**, 1020–1035. (doi:10.1016/j.physd.2011.02.012)
77. Scheffer M *et al.* 2009 Early-warning signals for critical transitions. *Nature* **461**, 53–59. (doi:10.1038/nature08227)

78. Lucarini V, Bódai T. 2017 Edge states in the climate system: exploring global instabilities and critical transitions. *Nonlinearity* **30**, R32–R66. (doi:10.1088/1361-6544/aa6b11)
79. Lucarini V, Bódai T. 2020 Global stability properties of the climate: melancholia states, invariant measures, and phase transitions. *Nonlinearity* **33**, R59–R92. (doi:10.1088/1361-6544/ab86cc)
80. Horsthemke W, Lefever R. 1984 *Noise-induced transitions: theory and applications in physics, chemistry, and biology*. Springer Series in Synergetics, vol. 15. Berlin, Germany: Springer.
81. Sornette D. 2006 Endogenous versus exogenous origins of crises. In *Extreme events in nature and society* (eds KHS Albeverio S, Jentsch V), pp. 95–119. Berlin, Germany.
82. Helmstetter A, Sornette D, Grasso J-R. 2003 Mainshocks are aftershocks of conditional foreshocks: how do foreshock statistical properties emerge from aftershock laws. *J. Geophys. Res. (Solid Earth)* **108**, 2046. (doi:10.1029/2002JB001991)
83. Sornette D. 2003 *Why stock markets crash (Critical Events in Complex Financial Systems)*. Princeton, NJ: Princeton University Press.
84. Sornette D, Helmstetter A. 2003 Endogenous versus exogenous shocks in systems with memory. *Physica A* **318**, 577–591. (doi:10.1016/S0378-4371(02)01371-7)
85. Wheatley S, Schatz M, Sornette D. 2018 The ARMA Point process and its estimation. (<http://arxiv.org/abs/1806.09948>).
86. Desai RC, Zwanzig R. 1978 Statistical mechanics of a nonlinear stochastic model. *J. Stat. Phys.* **19**, 1–24. (doi:10.1007/BF01020331)
87. Topaj D, Kye W-H, Pikovsky A. 2001 Transition to coherence in populations of coupled chaotic oscillators: a linear response approach. *Phys. Rev. Lett.* **87**, 074101. (doi:10.1103/PhysRevLett.87.074101)
88. Baek S-J, Ott E. 2004 Onset of synchronization in systems of globally coupled chaotic maps. *Phys. Rev. E* **69**, 066210. (doi:10.1103/PhysRevE.69.066210)
89. Bonilla LL, Casado J, Morillo M. 1987 Self-synchronization of populations of nonlinear oscillators in the thermodynamic limit. *J. Stat. Phys.* **48**, 571–591. (doi:10.1007/BF01019689)
90. Pecora LM. 1998 Synchronization conditions and desynchronizing patterns in coupled limit-cycle and chaotic systems. *Phys. Rev. E* **58**, 347–360. (doi:10.1103/PhysRevE.58.347)
91. Pecora LM, Carroll TL. 2015 Synchronization of chaotic systems. *Chaos* **25**, 097611. (doi:10.1063/1.4917383)
92. Graham R. 1977 Covariant formulation of non-equilibrium statistical thermodynamics. *Z. Phys. B* **26**, 397–405. (doi:10.1007/BF01570750)
93. Sznitman A. 1989 *Topics in propagation of chaos*, vol. 1464 of Hennequin PL. (eds) *Ecole d'Été de Probabilités de Saint-Flour XIX—1989*. Lecture Notes in Mathematics. Berlin, Germany: Springer.
94. Dawson DA, Gärtner J. 1987 Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* **20**, 247–308. (doi:10.1080/17442508708833446)
95. Bogachev VI, Krylov NV, Röckner M, Shaposhnikov SV. 2015 *Fokker-Planck-Kolmogorov equations*. Mathematical Surveys and Monographs, vol. 207. Providence, RI: American Mathematical Society.
96. Tamura Y. 1984 On asymptotic behaviors of the solution of a nonlinear diffusion equation. *J. Faculty Sci.* **31**, 195–221.
97. Bavaud F. 1991 Equilibrium properties of the Vlasov functional: the generalized Poisson–Boltzmann–Emden equation. *Rev. Mod. Phys.* **63**, 129–149. (doi:10.1103/RevModPhys.63.129)
98. Ogawa S, Yamaguchi YY. 2012 Linear response theory in the Vlasov equation for homogeneous and for inhomogeneous quasistationary states. *Phys. Rev. E* **85**, 061115. (doi:10.1103/PhysRevE.85.061115)
99. Patelli A, Ruffo S. 2014 General linear response formula for non integrable systems obeying the Vlasov equation. *Eur. Phys. J. D* **68**, 329. (doi:10.1140/epjd/e2014-50215-3)
100. Jackson JD. 1975 *Classical electrodynamics*, 2nd edn. New York, NY: Wiley.
101. Talebian E, Talebian M. 2013 A general review on the derivation of Clausius-Mossotti relation. *Optik* **124**, 2324–2326. (doi:10.1016/j.ijleo.2012.06.090)
102. Chekroun MD, Tantet A, Dijkstra HA, Neelin JD. 2020 Ruelle–Pollicott resonances of stochastic systems in reduced state space. Part I: theory. *J. Stat. Phys.* **179**, 1366–1402. (doi:10.1007/s10955-020-02535-x)
103. Shiino M. 1985 H-theorem and stability analysis for mean-field models of non-equilibrium phase transitions in stochastic systems. *Phys. Lett. A* **112**, 302–306. (doi:10.1016/0375-9601(85)90345-7)

104. Scheffer M. 2009 *Critical transitions in nature and society*. Princeton, NJ: Princeton University Press.
105. Kirkwood JG, Monroe E. 1941 Statistical mechanics of fusion. *J. Chem. Phys.* **9**, 514–526. (doi:10.1063/1.1750949)
106. Malrieu F. 2001 Logarithmic sobolev inequalities for some nonlinear pde's. *Stoch. Process. Appl.* **95**, 109–132. (doi:10.1016/S0304-4149(01)00095-3)
107. Frank T. 2013 Strongly nonlinear stochastic processes in physics and the life sciences. *ISRN Math. Phys.* **2013**, 1–28. (doi:10.1155/2013/149169)
108. Kloeden P, Platen E. 2011 *Numerical solution of stochastic differential equations*. Stochastic Modelling and Applied Probability. Berlin, Germany: Springer.
109. Lelièvre T, Nier F, Pavliotis GA. 2013 Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.* **152**, 237–274. (doi:10.1007/s10955-013-0769-x)
110. Duncan AB, Pavliotis GA, Zygalkis KC. 2017 Nonreversible Langevin samplers: splitting schemes, analysis and implementation. (<http://arxiv.org/abs/1701.04247>).
111. Duncan AB, Lelièvre T, Pavliotis GA. 2016 Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.* **163**, 457–491. (doi:10.1007/s10955-016-1491-2)